

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 18, 2005, 23:59:55 ; Search time 133.5 Seconds  
(without alignments)  
3702.470 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 1163  
Sequence: 1 atcgctgcctcgacgcg.....ataccaggtgaccatcaag 639

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10030225/runat\_18022005\_095900\_16266/app\_query.fasta\_1.775  
-DB=A\_Geneseq\_16Dec04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US10030225 @CGN 1.1 224 @runat\_18022005\_095900\_16266 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_16Dec04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1115	95.9	571	4 AAB49765	Aab49765 Human pro
2	1115	95.9	571	4 AAB88393	Aab88393 Human mem
3	430	37.0	81	4 AAM13752	Aam13752 Peptide #
4	430	37.0	81	4 ABB32685	Abb32685 Peptide #
5	430	37.0	81	4 AAM26150	Aam26150 Peptide #
6	430	37.0	81	4 AAB27526	Abb27526 Human pep
7	430	37.0	81	4 ABB18176	Abb18176 Protein #
8	430	37.0	81	4 AAM65885	Aam65885 Human bon
9	430	37.0	81	4 AAM53507	Aam53507 Human bra
10	430	37.0	81	4 ABG47540	Abg47540 Human liv

11	430	37.0	81	4 AAM01497	Aam01497 Peptide #
12	430	37.0	81	5 ABG35520	Abg35520 Human pep
13	230	19.8	173	7 ADF59409	Adf59409 Human pol
14	154	13.2	19938	6 ABP76679	Abp76679 Streptomy
15	148.5	12.8	546	4 ABUS3252	Abu53252 Human tes
16	148.5	12.8	1270	8 ADK67911	Adk67911 Human ext
17	148.5	12.8	1299	4 AAM24322	Aam24322 Human EST
18	148.5	12.8	1311	8 ADK67912	Adk67912 Human ext
19	148.5	12.8	1320	7 ADK65819	Adk65819 Angiogene
20	148.5	12.8	1404	2 AAR26049	Aar26049 MSF precu
21	148.5	12.8	1404	4 AAB29773	Aab29773 Human meg
22	148.5	12.8	1404	4 AAB60568	Aab60568 Human meg
23	148.5	12.8	1404	7 ADK65839	Adk65839 Angiogene
24	148.5	12.8	1404	8 ADM98014	Adm98014 Human meg
25	148.5	12.8	1415	4 AAU32262	Aau32262 Novel hum
26	148	12.7	513	4 ABUS3254	Abu53254 Human tes
27	140.5	12.1	551	4 ABUS3253	Abu53253 Human tes
28	135.5	11.7	377	5 ABP64957	Abp64957 Human pro
29	135	11.6	783	7 ABO69289	Abo69289 Pseudomon
30	134	11.5	399	3 AAB43375	Aab43375 Human ORF
31	133	11.4	339	7 ABO73785	Abo73785 Pseudomon
32	133	11.4	923	4 ABG08600	Abg08600 Novel hum
33	132	11.3	246	7 ABO71543	Abo71543 Pseudomon
C	130.5	11.1	2655	7 ADO59401	Ado59401 Anthracia
35	130	11.2	489	4 ABG14285	Abg14285 Novel hum
36	129	11.1	1518	6 ABJ18375	Abj18375 Breast sp
37	128.5	11.0	386	7 ABO75475	Abo75475 Pseudomon
38	128	11.0	373	7 ABO80173	Abo80173 Pseudomon
39	128	11.0	410	7 ABO77066	Abo77066 Pseudomon
40	128	11.0	554	4 ABG14613	Abg14613 Novel hum
41	128	11.0	655	8 ABM84305	Abm84305 Human dia
42	127.5	11.0	223	8 ADS30809	Ads30809 Bacterial
43	127.5	11.0	233	7 ABO84024	Abo84024 Pseudomon
44	127.5	11.0	1953	6 AAE36881	Aae36881 Nephila m
45	127	10.9	806	8 ADO17018	Ado17018 Predicted

ALIGNMENTS

RESULT 1

AAB49765  
ID AAB49765 standard; protein; 571 AA.  
XX

AC AAB49765;

DT 20-APR-2001 (first entry)

XX Human proliferation differentiation factor amino acid sequence.

XX Human; proliferation differentiation factor; haematopoietic function.

OS Homo sapiens.

XX WO200104312-A1.

XX 18-JAN-2001.

XX 06-JUL-2000; 2000WO-JP004514.

XX 08-JUL-1999; 99JP-00194179.

XX 18-OCT-1999; 99US-0159586P.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;

XX WPI; 2001-138354/14.

XX N-PSDB; AAF29348.

XX Proliferation-differentiation factor protein encoded by PSEC137 cloned from human cDNA library, being hematopoietic factor for inducing differentiation of blood cells, used to maintain hematopoietic function.













3









**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 19, 2005, 00:09:26 ; Search time 31.5 Seconds  
(without alignments)  
3028.619 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 1163

Sequence: 1 atcgctgcctccgcagcgcg.....ataaccaggtgaccatcaag 639

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Dgapop 6.0	Dgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPO.spool\_P/US10030225/runat\_18022005\_095901\_16296/app\_query.fasta\_1.775  
-DB=Issued\_Patents\_AA-QPM-fastan-SUFFIX=rai-MINMATCH=0.1-LOOPCL=0  
-LOOPEXT=0-UNITS=bits-START=1-END=1-MATRIX=blosum62-TRANS-human40.cdi  
-LIST=45-DOCALIGN=200-THR SCORE=pct-THR MAX=100-THR MIN=0-ALIGN=15  
-MODE=LOCAL-OUTFMT=pto-NORM=ext-HEAPSIZE=500-MINLEN=0-MAXLEN=2000000000  
-USER=US10030225@cgn 1.1.46@runat\_18022005\_095901\_16296-NCPU=6-ICPU=3  
-NO MMAP-LARGEQUERY-NEG SCORES=0-WAIT-DSPBLOCK=100-LOGLOG  
-DEV TIMEOUT=120-WARN TIMEOUT=30-THREADS=1-XGAPOP=10-XGAPEXT=0.5-FGAPOP=6  
-FGAPEXT=7-YGAPOP=10-YGAPEXT=0.5-DELOP=6-DELEXT=7

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148.5	12.8	941	4	US-07-757-022B-14
2	148.5	12.8	1022	4	US-07-757-022B-84
3	148.5	12.8	1038	4	US-07-757-022B-74
4	148.5	12.8	1049	4	US-07-757-022B-58
5	148.5	12.8	1140	4	US-07-757-022B-104
6	148.5	12.8	1270	4	US-07-757-022B-44
7	148.5	12.8	1311	4	US-07-757-022B-42
8	148.5	12.8	1313	4	US-07-757-022B-142
9	148.5	12.8	1314	4	US-07-757-022B-50
10	148.5	12.8	1320	4	US-07-757-022B-46
11	148.5	12.8	1320	4	US-07-757-022B-60
12	148.5	12.8	1320	4	US-10-164-595-58

Sequence 48, Appl  
Sequence 40, Appl  
Sequence 52, Appl  
Sequence 2, Appl  
Sequence 62, Appl  
Sequence 1, Appl  
Sequence 18035, A  
Sequence 18027, A  
Sequence 2819, A  
Sequence 2812, A  
Sequence 32770, A  
Sequence 10, Appl  
Sequence 1, Appl  
Sequence 161, App  
Sequence 47, Appl  
Sequence 47, Appl  
Sequence 19845, A  
Sequence 26099, A  
Sequence 21753, A  
Sequence 8970, Ap  
Sequence 26135, A  
Sequence 31529, A  
Patent No. 5168049  
Patent No. 5168049  
Sequence 21753, A  
Sequence 11715, A  
Sequence 47, Appl  
Sequence 11716, A  
Sequence 30843, A

#### ALIGNMENTS

#### RESULT 1

US-07-757-022B-14  
; Sequence 14, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 941 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-14

Alignment Scores:
Pred. No.: 9,19e-05 Length: 941
Score: 148.50 Matches: 64
Percent Similarity: 32.48% Conservative: 12
Best Local Similarity: 27.35% Mismatches: 124
Query Match: 12.77% Indels: 34
DB: 4 Gaps: 5

US-10-030-225-1_COPY_58_696 (1-639) x US-07-757-022B-14 (1-941)
QY 18 CCGAGCCGGGCTCTCTCTGCTGCTGCTGCGCGCTGCTGCGAGCGGCGCTAGG 77
Db 299 ProAlaProThrThrProLysGluProAlaProThrThrThrLysGluProSerProThr 318
QY 78 GCTCCCGCTGAGAGGCC-----CGG 98
Db 319 ThrProLysGluProAlaProThrThrThrLysSerAlaProThrThrThrLysGluPro 338
QY 99 GCTCCGCGGACACACCGGCTGGGAGCTCACAGGCTCGCAGAGGCTCTCA-----GCCTC 152
Db 339 AlaProThrThrThrLysSerAlaPro-ThrThrProLysGluProSerProThrThrTh 358
QY 153 CCCAGATCTAGGCTCTGAAGAGAGAGAGAGGAGGACACACTGCTCCCGACACCCACT 212
Db 358 rlysgluProAlaProThrThrProLysGluProAlaProThrThrProLysLysProAl 378
QY 213 GCAGGACAGGACACACACATGGATGCTGACTGCTACTGAGCAGCAGCAGCAGCCATGACC-- 270
Db 378 aProThrThrProLysGluProAlaProThrThrProLysGluProAlaProThrThrTh 398
QY 271 -----CCAGGCAACACACCCCTCCCGAGCAGCC 299
Db 398 rlysglyProAlaProThrAlaProLysGluProAlaProThrThrProLysGluThrAl 418
QY 300 AGAGGTTACTCGGTGCGGCTGAGCTGAGAGTGCCTGGGATGTCGAGCAGACACCTT 359
Db 418 aProThrThrProLysLysLeuThrProThrThrProLysLysLeuAlaProThrThrPr 438
QY 360 GAGTACCCCTAACCTGTATACCAGGCTTCAGCCTCCCA-----GATCCTAG 407
Db 438 oGluLysProAlaProThrThrProGluLysLeuAlaProThrThrProGluLysProTh 458
QY 408 GCCTCTGAGGGAAGAGGAGGAGGACGACTGCTCCCGAGACCCACCTGCGAGGAGAGCT 467
Db 458 rProThrThrProGluLysProAlaProThrThrProLysAlaAlaAlaProAsnThrPr 478
QY 468 ACACCAACATGGATGTGACTGTCACTGAGCAGCAGCCCTGACC-----CCAGG 518
Db 478 oLysGluProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAl 498
QY 519 GAATGCCACGCTCCCGAGGACCCAGGAGGTACTCTGCTGCTGCTGCTGCTGCTGCTGCT 578
Db 498 aProThrThrProLysGluThrAlaProThrThrProLysGlyThrAlaProThrThrle 518
QY 579 GCCAGATTGGTCCAGCGAACCTTGAGTACCCTTAACCT 618
```

```
Db 518 uLysGluProAlaProThrThrProLysLysProAlaPro 531
RESULT 2
US-07-757-022B-84
; Sequence 84, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-84

Alignment Scores:
Pred. No.: 9,35e-05 Length: 1022
Score: 148.50 Matches: 64
Percent Similarity: 32.48% Conservative: 12
Best Local Similarity: 27.35% Mismatches: 124
Query Match: 12.77% Indels: 34
DB: 4 Gaps: 5

US-10-030-225-1_COPY_58_696 (1-639) x US-07-757-022B-84 (1-1022)
QY 18 CCGAGCCGGGCTCTCTCTGCTGCTGCTGCGCGCTGCTGCGAGCGGCGCTAGG 77
Db 380 ProAlaProThrThrProLysGluProAlaProThrThrThrLysGluProSerProThr 399
QY 78 GCTCCCGCTGAGAGGCC-----CGG 98
```

400 ThrProLysGluProAlaProThrThrThrLysSerAlaProThrThrThrLysGluPro 415

99 GCTCCGCGACACACGCGCTGGGAGCCTCACGAGGCTCGCAGAGGTCTCA-----GCCTC 152

420 AlaProThrThrThrLysSerAlaProThrThrProLysGluProSerProThrThrTh 439

153 CCAGATCTTAGCCTCTGAAGAAAGGAGGAGGACCACTGCTCCCGAAGAACCCACT 212

439 rLysGluProAlaProThrThrProLysGluProAlaProThrThrProLysLysProAl 459

213 GCAGGCAGAGCCACACCACTGGATGCTGACCTGCTCCTGAGCCAGCAGCCATGACC-- 270

459 aProThrThrProLysGluProAlaProThrThrProLysGluProAlaProThrThrTh 479

271 -----CCAGCAACACACACCCCTCCAGGACCCCTCCAGGACCCCT 299

479 rLysLysProAlaProThrAlaProLysGluProAlaProThrThrProLysGluThrAl 499

300 AGAGTTACTTCGTTGGCTGGAGCTGCAGAACTCCGGGATTGGCCAGCACAACCTT 359

499 aProThrThrProLysLysLeuThrProThrThrProLysLysLeuAlaProThrThrPr 519

360 GAGTACCCCTAACCTGATACCCAGGCTTCAGCCTCCCA-----GATCCTAG 407

519 oGluLysProAlaProThrThrProLysGluLeuAlaProThrThrProLysGluProTh 539

408 GCCTCTGAGGGAAGAGGAGGAGGACGACCTGCTCCCGAAGAACCCACTGCGAGCAGAGCT 467

539 rProThrThrProLysGluProAlaProThrThrProLysAlaAlaAlaProAlaThrPr 559

468 ACACCAACATGATTTGAGCTGTCACCTGAGCCAGCAGCCCTGACC-----CCAGG 518

559 oLysGluProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAl 579

519 GAATGCCAGCCTCCAGGACCCAGGAGTTACTTCCCTGCTGCTGAGCTGAGAGCT 578

579 aProThrThrProLysGluThrAlaProThrThrProLysGlyThrAlaProThrThrLe 599

579 GCCAGAAATGGTCCACGCAACTTGAGTACCCCTTAACCT 618

599 uLysGluProAlaProThrThrProLysLysProAlaPro 612

RESULT 3

US-07-757-022B-74

; Sequence 74, Application US/07757022B

; Patent No. 6433142

; GENERAL INFORMATION:

; APPLICANT: Geener, Thomas G.

; APPLICANT: Clark, Stephen C.

; APPLICANT: Turner, Katherine

; APPLICANT: Hewick, Rodney M.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/757,022B

; FILING DATE: 19910910

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; PRIOR APPLICATION DATA:

407	ProAlaProThrThrProLysGluProAlaProThrThrThrLysGluProSerProThr	426
78	GCTCCCGGTGAAGAAGCC	
427	ThrProLysGluProAlaProThrThrThrLysSerAlaProThrThrThrLysGluPro	446
99	GCTCCGGGACACGCGCTGGGAGCTCACGAGGCTCGCAGAGCTCA	152
447	AlaProThrThrThrLysSerAlaProThrThrProLysGluProSerProThrThrTh	466
153	CCACAGTCCTAGCGCTCTGAAGAGAGGAGGAGGCACCACTGCTCCCGACAACCCACCT	212
466	LysGluProAlaProThrThrProLysGluProAlaProThrThrProLysLysProAl	486
213	GCAGGACAGCCACCAACATGGATGCTGGACTGTCACTGAGCCAGCAGCCATGACC	270
486	aProThrThrProLysGluProAlaProThrThrProLysGluProAlaProThrThrTh	506
271	-----CCAGGCAACACCAACCCCTCCCGAGCAGCC	299
506	LysLysProAlaProThrAlaProLysGluProAlaProThrThrProLysGluThrAl	526
300	AGAGTTACTCGTTCGGCTGAGCTGCAGAGCTCCGGGATGTGCCAGACACACCTT	359
526	aProThrThrProLysLysLeuThrProThrThrProGluLysLeuAlaProThrThrPr	546
360	GAGTACCCCTAACCTGATACCCAGGCTTCAGCTCTCCCA	407
546	oGluLysProAlaProThrThrProGluGluLeuAlaProThrThrProGluGluProTh	566
408	GCCTCTGAGGAAGAGAGGAGGACGACTGCTCCCGAGAACCCACCTCGCAGCAGAGCT	467
566	rProThrThrProGluGluProAlaProThrThrProLysAlaAlaAlaProAsnThrPr	586
468	ACACCAACATGGATGTGGACTGTCACTGAGCCAGCAGCCCTGACC	518
586	oLysGluProAlaProThrThrProLysGluProAlaProThrThrProLysGluProAl	606
519	GAATGCCAGGCTCCAGGACCCAGGAGGTACTCCCTTGCTGCTGGAGCTGCAGAGCT	578
606	aProThrThrProLysGluThrAlaProThrThrProLysGlyThrAlaProThrThrLe	626
579	GCCAGAAATGGTCCACGCAACCTTGATGATCCCTTAACCTT	618
626	uLysGluProAlaProThrThrProLysLysProAlaPro	639

FILING DATE: 08-AUG-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ceert, Luann  
 REGISTRATION NUMBER: 31,822  
 REFERENCE/DOCKET NUMBER: GI 5190  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)876-1170  
 TELEFAX: (617)876-5851  
 INFORMATION FOR SEQ ID NO: 58:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1049 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-757-022B-58

Alignment Scores:	9.4e-05	1049
Pred. No.:	148.50	64
Score:	32.48%	12
Percent Similarity:	32.48%	124
Best Local Similarity:	27.95%	34
Query Match:	12.77%	5
DB:	4	

US-10-030-225-1\_COPY\_58\_696 (1-639) x US-07-757-022B-58 (1-1049)

US-10-030-225-1 COPY 58 696 (1-639) X US-07-757-022B-58 (1-1049)

18 CCGAGCCGGGCTCCTCCTCTGCGTGCTGCTGGCGGCTGCTGGAGCGCGCTAGG 77

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-104

Alignment Scores:
Pred. No.: 1 9,578-05 Length: 1140
Score: 148.50 Matches: 64
Percent Similarity: 32.48% Conservative: 12
Best Local Similarity: 27.35% Mismatches: 124
Query Match: 12.77% Indels: 34
DB: 4 Gaps: 5

US-10-030-225-1_COPY_58_696 (1-639) x US-07-757-022B-104 (1-1140)

QY 18 CCGAGCGGGCTCCCTCTCGTCTGCTCTGCGCGCGCTGCTGGAGCGCGCTAGG 77
|||::: ||| |||
Db 498 ProAlaProThrThrProLysGluProAlaProThrThrThrLysGluProSerProThr 517
|||::: ||| |||

QY 78 GTCCTCCGTTGAAGAGCC-----GCG 98
|||::: ||| |||
Db 518 ThrProLysGluProAlaProThrThrThrLysSerAlaProThrThrThrLysGluPro 537
|||::: ||| |||

QY 99 GTCCTCGCGGACACCGCGCTGGAGGCTCACAGAGCTCGAGAGTCTCA-----GCCTC 152
|||::: ||| ||| ||| |||
Db 538 AlaProThrThrThrLysSerAlaPro-ThrThrProLysGluProSerProThrThrTh 557
|||::: ||| ||| ||| |||

QY 153 CCCAGATCTTAGCGCTCTGAAGGAAGAGGAGGAGGCACCACTGCTCCCCAGAACCCACCT 212
: : :||| ||| ||| ||| |||:::
Db 557 rLysGluProAlaProThrThrThrProLysGluProAlaProThrThrProLysLysProAl 577
|||::: ||| ||| ||| ||| |||

QY 213 GCAGGCAGAGCCACACCAACATGGATGCTGACTGTCTCACTGAGCCAGCAGCCATGACC-- 270
|||::: ||| ||| ||| ||| |||
Db 577 aProThrThrProLysGluProAlaProThrThrThrProLysGluProAlaProThrThrTh 597
|||::: ||| ||| ||| ||| |||

QY 271 -----CAGGCAACACACACCCCTCCAGGACCCC 299
||| ||| ||| ||| ||| ||| |||
Db 597 rLysLysProAlaProThrAlaProLysGluProAlaProThrThrProLysGluThrAl 617
||| ||| ||| ||| ||| ||| |||

QY 300 ACAGGTTACTCGGTTGCGCGTGGAGCTGCAGAACTGCCGGGATGGCCAGCAGCAACCTT 359
||| ||| ||| ||| ||| ||| ||| |||
Db 617 aProThrThrProLysLysLeuThrProThrThrProGluLeuAlaProThrThrThrThrPr 637
||| ||| ||| ||| ||| ||| ||| |||

QY 360 GAGTACCCCTAACCCCTGATACCAGGCTTCAGCCTCCCCA-----GATCCTAG 407
||| ||| ||| ||| ||| ||| ||| |||
Db 637 oGluLysProAlaProThrThrThrProGluLeuAlaProThrThrThrProGluLysLeuAlaProThr 657
||| ||| ||| ||| ||| ||| ||| |||

QY 408 GCCTCTGAGGGAAGAGGAGGAGGCACGACTGCTCCCGAGAACCCACCTCGCAGCAGAGCT 467
||| ||| ||| ||| ||| ||| ||| |||
Db 657 rProThrThrProGluGluProAlaProThrThrThrProLysLeuAlaProAlaProThrPr 677
||| ||| ||| ||| ||| ||| ||| |||

```









US-07-757-022B-60  
; Sequence No. 6433142  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Geaner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte  
; NUMBER OF SEQUENCES: 143







Db 676 uLysGluProAlaProThrThrProLysLysProAlaPro 689

Search completed: February 19, 2005, 00:24:15  
Job time : 45.5 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 19, 2005, 00:21:47 ; Search time 118.5 Seconds  
(without alignments)  
3529.232 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 1163  
Sequence: 1 atcgctgcgtccgcgcgcg.....ataaccaggtagaccatcaag 639

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 2760536

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p,model -DEV=xlp  
-Q=/cgn2\_1/USTO\_spool\_p/US10030225/runat\_18022005\_095903\_16366/app\_query.fasta\_1.775  
-DB=Published Applications\_AA -QFMT=fastan -SUFFIX=rapb -MINWATCH=0.1  
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-NCPUE=6 -ICPU=3 -NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100  
-LONLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pcp:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pcp:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pcp:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pcp:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pcp:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pcp:\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp:\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	1053.5	90.6	658	11	US-09-939-853A-94	Sequence 94, Appl
2	879	75.6	538	15	US-10-074-978A-403	Sequence 403, App
3	430	37.0	81	9	US-09-864-761-33474	Sequence 33474, A
4	154.5	13.3	410	16	US-10-437-963-196665	Sequence 196665,
5	148.5	12.8	941	13	US-10-124-557-14	Sequence 14, Appl
6	148.5	12.8	1022	13	US-10-124-557-74	Sequence 84, Appl
7	148.5	12.8	1038	13	US-10-124-557-74	Sequence 74, Appl
8	148.5	12.8	1049	13	US-10-124-557-58	Sequence 58, Appl
9	148.5	12.8	1140	13	US-10-124-557-104	Sequence 104, Appl
10	148.5	12.8	1270	13	US-10-124-557-44	Sequence 44, Appl
11	148.5	12.8	1311	13	US-10-124-557-42	Sequence 42, Appl
12	148.5	12.8	1313	13	US-10-124-557-142	Sequence 142, App
13	148.5	12.8	1314	13	US-10-124-557-50	Sequence 50, Appl
14	148.5	12.8	1320	13	US-10-124-557-46	Sequence 46, Appl
15	148.5	12.8	1320	13	US-10-124-557-60	Sequence 60, Appl
16	148.5	12.8	1354	13	US-10-124-557-48	Sequence 48, Appl
17	148.5	12.8	1361	13	US-10-124-557-40	Sequence 40, Appl
18	148.5	12.8	1363	13	US-10-124-557-52	Sequence 52, Appl
19	148.5	12.8	1404	9	US-09-802-207-30	Sequence 30, Appl
20	148.5	12.8	1404	11	US-09-897-188-1	Sequence 1, Appli
21	148.5	12.8	1404	13	US-10-124-557-2	Sequence 2, Appli
22	148.5	12.8	1404	13	US-10-124-557-62	Sequence 62, Appl
23	148.5	12.8	1404	17	US-10-868-577A-55	Sequence 55, Appl
24	145	12.5	404	16	US-10-437-963-116565	Sequence 116565,
25	144.5	12.4	192	16	US-10-437-963-171117	Sequence 171117,
26	144.5	12.4	625	16	US-10-437-963-165015	Sequence 165015,
27	144.5	12.4	19723	15	US-10-084-846A-5	Sequence 5, Appli
28	139.5	12.0	271	16	US-10-437-963-181500	Sequence 181500,
29	136.5	11.7	481	16	US-10-437-963-176449	Sequence 176449,
30	134	11.4	420	16	US-10-437-963-201378	Sequence 201378,
31	134	11.5	655	16	US-10-437-963-141788	Sequence 141788,
32	133.5	11.5	19662	15	US-10-084-846A-6	Sequence 6, Appli
33	132.5	11.4	19608	15	US-10-084-846A-8	Sequence 8, Appli
34	131	11.3	537	16	US-10-437-963-188969	Sequence 188969,
35	130	11.1	469	15	US-10-425-114-43328	Sequence 43328, A
36	129.5	11.1	366	16	US-10-437-963-197924	Sequence 197924,
37	129	11.1	205	16	US-10-437-963-190742	Sequence 190742,
38	129	11.1	550	16	US-10-437-963-145800	Sequence 145800,
39	129	11.1	1518	11	US-09-989-890-184	Sequence 184, App
40	128.5	11.0	423	16	US-10-437-963-158031	Sequence 158031,
41	128.5	11.0	466	16	US-10-437-963-195119	Sequence 195119,
42	128.5	11.0	582	14	US-10-156-761-11938	Sequence 11938, A
43	127.5	11.0	223	17	US-10-369-493-19842	Sequence 19842, A
44	127.5	11.0	1953	17	US-10-488-056-42	Sequence 42, Appl
45	127	10.9	19652	15	US-10-084-846A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-09-939-853A-94  
; Sequence 94, Application US/09939853A  
; Publication No. US20040039163A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess et al.  
; TITLE OF INVENTION: No. US20040039163A1el Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-099  
; CURRENT APPLICATION NUMBER: US/09/939, 853A  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR FILING DATE: 2001-08-27  
; PRIOR FILING DATE: 2000-08-25  
; PRIOR APPLICATION NUMBER: 60/267,300  
; PRIOR FILING DATE: 2001-02-08  
; PRIOR APPLICATION NUMBER: 60/269,961  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/277,337  
; PRIOR FILING DATE: 2001-03-20  
; NUMBER OF SEQ ID NOS: 159  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 94  
; LENGTH: 658





```

, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/006669
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/006665
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/006668
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/006663
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/006662
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/006661
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: PCT/US01/006670
, PRIOR FILING DATE: 2001-01-30
, PRIOR APPLICATION NUMBER: US 60/234,687
, PRIOR FILING DATE: 2000-09-21
, PRIOR APPLICATION NUMBER: US 09/608,408
, PRIOR FILING DATE: 2000-06-30
, PRIOR APPLICATION NUMBER: US 09/774,203
, PRIOR FILING DATE: 2001-01-29
, NUMBER OF SEQ ID NOS: 49117
, SEQ ID NO 33474
, LENGTH: 81
, TYPE: PRT
, ORGANISM: Homo sapiens
, FEATURE:
, OTHER INFORMATION: MAP TO AF111168.2
, OTHER INFORMATION: EXPRESSED IN PLACENT
, OTHER INFORMATION: EXPRESSED IN BT474,
, OTHER INFORMATION: EXPRESSED IN BONE MAR
, OTHER INFORMATION: EXPRESSED IN HEL100,
, OTHER INFORMATION: EXPRESSED IN FETAL L
, OTHER INFORMATION: EXPRESSED IN LUNG, S
, OTHER INFORMATION: EXPRESSED IN ADULT L
, OTHER INFORMATION: EXPRESSED IN BRAIN, S
, OTHER INFORMATION: EXPRESSED IN HELA, S
, OTHER INFORMATION: EXPRESSED IN HEART, S
, OTHER INFORMATION: EST HUMAN HIT: BE281
, OTHER INFORMATION: EST HUMAN HIT: BE731
, OTHER INFORMATION: SWISSPROT HIT: Q0281
, US-09-864-761-33474

```

<b>Alignment Scores:</b>					
Pred. No.:	9.34e-24	Length:	81		
Score:	430.00	Matches:	79		
Percent Similarity:	98.77%	Conservative:	1		
Best Local Similarity:	97.53%	Mismatches:	0		
Query Match:	36.97%	Indels:	0		
DB:	9	Gaps:	0		
US-10-030-225-1_COPY_58_696 (1-639) x US-09-864-761-33474 (1-81)					
QY	142	GTCTCAGCCTCCCGAGATCCTTAGGCCTCTGAAGAAGGAGGAGGCCACCACTGCTCCCC	201		
Db	1	VaIserAlaserProAspPProArgProLeuysGlUGluGluAlaProLeuLeuPro	20		
QY	202	AGAACCCACCTGCAGGCAGGCCACACCACATGGATGCTGGACTGTCACTGAGCCAGCA	261		
Db	21	ArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAla	40		
QY	262	GCCATGACCCGAGCAACACCACCCCTCCAGGACCCCAGAGTTTACTTCGGTTCGGGCTG	321		
Db	41	AlaMetThrProGlyAsnAlaThrProProArgThrProGluValThrProLeuArgLeu	60		
QY	322	GAGCTGCAGAAAGCTGCCGGGATTGGCCAGCACACCTTGAGTACCCCTTAACCCCTGATACC	381		
Db	61	GluLeuGlnLysLeuProGlyLeuAlaAsnThrThrLeuSerThrProAsnProAspThr	80		
QY	382	CAG 384			
Db	81				
Db	81				

```

US-10-030-225-1_COPY_58_696 (1-639) x US-10-437-963-196665 (1-410)

QY 9 GCTCCGACGAGCGGGCTCTCTCTGGGTGCTGCTGG----- 52
Db 9 AlaProLeuProThraAlaSerProProAlaAlaArgArgGlnProThraAlaVal 109
QY 53 CGCGCTCTGAGGCGGGCTAGGGTCCCGCTGAAGACCGCGGCTCCGCGGACCAC 112
Db 53 AlaProLeuProThraAlaSerProProAlaAlaArgArgGlnProThraAlaVal 109
QY 110 HisArgGtrPArgArg---SerGlnProValArgProArg-AlaProLeuThraAl 128
Db 110 HisArgGtrPArgArg---SerGlnProValArgProArg-AlaProLeuThraAl 128
QY 113 GSCCTGGGAGCTCACGAGGCTCGCAGGCTCTCAGCCTCCCGATCTCAGGCTCTGA 172
Db 113 GSCCTGGGAGCTCACGAGGCTCGCAGGCTCTCAGCCTCCCGATCTCAGGCTCTGA 172
QY 128 aSerProProSerSerAlaAlaArg-----AlaAlaAl 140
Db 128 aSerProProSerSerAlaAlaArg-----AlaAlaAl 140
QY 173 AGGAAGAGGAGGAGCACCAGTGTCTCCCGAACCCACTGTGAGGAG- -GCCACACC 229
Db 173 AGGAAGAGGAGGAGCACCAGTGTCTCCCGAACCCACTGTGAGGAG- -GCCACACC 229
QY 140 aHisArgGlnSerGly---ProAlaArgSerProProAlaHisArgGlnSerAlaPr 159
Db 140 aHisArgGlnSerGly---ProAlaArgSerProProAlaHisArgGlnSerAlaPr 159
QY 230 AACATGGATGCTGGACTGTCTACTGAGCCAGCAGCATGATCCCGAGCAACACCCCTC 289
Db 230 AACATGGATGCTGGACTGTCTACTGAGCCAGCAGCATGATCCCGAGCAACACCCCTC 289
QY 159 othr-----ArgArgSerArgSerProSerProHisProThr 172
Db 159 othr-----ArgArgSerArgSerProSerProHisProThr 172
QY 290 CCAGGACCCAGAGGTTACTCGCTTGGCTGGAGCTGCAGAGCTGCCGGGATGGCCA 349
Db 290 CCAGGACCCAGAGGTTACTCGCTTGGCTGGAGCTGCAGAGCTGCCGGGATGGCCA 349
QY 172 rHisProProArgGly-AlaPro-CysSerProGlyGluAspSerCys---ThrTrpPro 190
Db 172 rHisProProArgGly-AlaPro-CysSerProGlyGluAspSerCys---ThrTrpPro 190
QY 350 GCACACCTTGATACCCCTTAACCTGATACCCAGGCTTCAGCCTCCCGAGATCCTAGGC 409
Db 350 GCACACCTTGATACCCCTTAACCTGATACCCAGGCTTCAGCCTCCCGAGATCCTAGGC 409
QY 191 Ser-----ProCysThrProLeuProArgHisgln----- 200
Db 191 Ser-----ProCysThrProLeuProArgHisgln----- 200
QY 410 CTCTGAGGGAAGAGGAGGAGCAGACTGTCTCCCGAGACCCACCTTCGAGGACAGCTAC 469
Db 410 CTCTGAGGGAAGAGGAGGAGCAGACTGTCTCCCGAGACCCACCTTCGAGGACAGCTAC 469
QY 201 -----ArgArgArgHisGlyCysAlaProAlaProSer-AlaAlaAlaAlaTh 216
Db 201 -----ArgArgArgHisGlyCysAlaProAlaProSer-AlaAlaAlaAlaTh 216

```

Alignment Scores:		
Pred. No.:	0.00928	941
Score:	148.50	64
Percent Similarity:	32.48%	12
Best Local Similarity:	27.35%	Conservative: 124
Query Match:	12.77%	Mismatches: 24
DB:	13	Indels: 5
		Gaps: 5





**Alignment Scores:**

```

/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02140
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/124,557
/ FILING DATE: 16-Apr-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/643,502
/ FILING DATE: 18-JAN-1991
/ APPLICATION NUMBER: US 07/546,114
/ FILING DATE: 29-JUN-1990
/ APPLICATION NUMBER: US 07/457,196
/ FILING DATE: 29-DEC-1989
/ APPLICATION NUMBER: US 07/390,901
/ FILING DATE: 08-AUG-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Cserr, Luann
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: G1 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 876-1170
/ TELEFAX: (617) 876-5851
/ INFORMATION FOR SEQ ID NO: 44:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1270 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 44:
/ US-10-124-557-44

Alignment Scores:
Pred. No.: 0.00949 Length: 1270
Score: 148.50 Matches: 64
Percent Similarity: 32.48% Conservative: 124
Best Local Similarity: 27.35% Mismatches: 12
Query Match: 12.77% Indels: 34
DB: 13 Gaps: 5

US-10-030-225-1_COPY_58_696 (1-639) x US-10-124-557-44 (1-12)
Qy 18 CCGAGCGGGCTCTCTCTCTGGTCTCTCTCTGGCGGCGTGTGTG
|||::| |||
Db 364 ProAlaProThrThrProLysGluProAlaProThrThrLysGluPro
|||::| |||
Qy 78 GCTCCCGGTGAGAGCC-----
|||::| |||
Db 384 ThrProLysGluProAlaProThrThrThrLysSerAlaProThrThr
|||::| |||
Qy 99 GCTCCGCGGACCCAGCGCTGGGAGCCTCAGAGGCTCGCAGAGGCTCT
|||::| |||
Db 404 AlaProThrThrThrLysSerAlaPro-ThrThrProLysGluPro
|||::| |||
Qy 153 CCCAGATCTAGGCTCTTGAGAGGAGGAGGAGGCCACCACTGCTGTC
:::| |||
Db 423 rlyysGluProAlaProThrThrProLysGluProAlaProThrThr
:::| |||
Qy 213 GCAGGCGAGGCCACACCAACATGGATGCTGAGCTGTCACTGAGGCCA
|||::| |||
Db 443 aProThrThrProLysGluProAlaProThrThrProLysGluPro
|||::| |||
Qy 271 -----CCAGGCAACACACAC
|||
Db 463 rlyysLysProAlaProThrAlaProLysGluProAlaProThrThr
|||
Qy 300 AGAGGTTACTCGGTGGCGGTGGAGCTGCAGAACTGCCGGGATGTG
|||||::| |||

```

404	AlaProThrThrLysSerAlaPro-ThrThrProLysGluProSerProThrThrTh	423
153	CCGAGATCTTAGCCCTCTGAAGGAGAGGAGGAGGACCACTGCTCCCCAGAACCCACCT	212
423	rLysGluProAlaProThrThrProLysGluProAlaProThrThrProLysLysProAl	443
213	GCAGGCAGAGCCACACCAATGATGCTGGACTGTCTCACTGAGCCAGCAGCCATGACC--	270
443	aProThrThrProLysGluProAlaProThrThrProLysGluProAlaProThrThrTh	463
271	-----CCAGGCAACACCACCCCTCTCCAGAGCCCC	299
463	rLysLysProAlaProThrAlaProLysGluProAlaProThrThrProLysGluThrAl	483
300	AGAGGTTACTCCGTTGGCGTGGAGCTGCAGAGCTCCCGGATGGCCAGACCAACCTT	359







## RESULT 14



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: February 19, 2005, 00:09:01 ; Search time 37 Seconds  
(without alignments)  
3323.379 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 1163

Sequence: 1 atgcgtgcctccgcgacgcg.....ataaccaggtgaccatcaag 639

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPO\_001225/runat\_18022005\_095901\_16284/app\_query.fasta\_1.775  
-DB=PIR\_79 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -WATRIX=Blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US1010225\_QCGN 1 1 63 @runat\_18022005\_095901\_16284 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150.5	12.9	460	2	T33110
2	136.5	11.7	880	2	D89756
3	134	11.5	653	2	E84682
4	133.5	11.5	383	2	S32975
5	133.5	11.4	2639	2	T31328
6	128	11.0	395	2	H75457
7	127.5	11.0	294	2	A37232
8	126.5	10.9	825	1	EDBEXD
9	126	10.8	377	2	A48018
10	126	10.8	1428	2	T08852
11	126	10.8	2649	2	T51023
12	125	10.7	3149	1	Q8BEH8
13	123.5	10.5	446	2	T45525
14	123.5	10.6	1323	2	I78557

15	123.5	10.6	1356	1	C45219
16	122	10.5	627	2	A44112
17	121.5	10.4	839	2	F75518
18	121	10.4	2688	2	I49477
19	120	10.3	3938	2	T42761
20	119.5	10.3	775	1	EDBE11
21	119	10.2	803	2	F59433
22	118.5	10.2	1133	2	A54164
23	118	10.1	801	2	T29018
24	117.5	10.1	1224	2	T14007
25	117.5	10.1	1585	2	T31611
26	117.5	10.1	2715	2	T13049
27	117	10.1	580	2	T43481
28	116.5	10.0	216	2	F51920
29	116.5	10.0	283	2	S35332
30	116.5	10.0	440	2	JC7807
31	116.5	10.0	846	2	T21700
32	116.5	9.9	967	2	S66852
33	116.5	10.0	1241	2	T18311
34	116	10.0	291	2	S27721
35	116	10.0	1186	2	T19050
36	116	10.0	3020	2	A43932
37	115.5	9.9	521	1	VGBEHB
38	115.5	9.9	2187	2	T30826
39	115	9.9	699	2	C43674
40	114	9.8	1334	2	T50568
41	114	9.7	13288	2	T03099
42	113.5	9.8	530	2	A45690
43	113.5	9.8	698	2	T17261
44	113.5	9.8	1952	2	T48814
45	112.5	9.7	676	1	EDBE23

ALIGNMENTS

RESULT: 1

T33110

hypothetical protein C18H7.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T33110

R:Tin-Mollam, A.; Fronick, W.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid C18H7.

A:Reference number: Z21284

A:Accession: T33110

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-460 <TIN>

A:Cross-references: UNIPROT:Q9GZF7; EMBL:AF067607; PIDN:AAC17641.1; GSPDB:GN00022; CESP:

C:Experimental source: strain Bristol N2; clone C18H7

C:Genetics:

A:Gene: CESP:C18H7.3

A:Map position: 4

A:Introns: 84/1

C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Alignment Scores:			
Pred. No.:	0.00197	Length:	460
Score:	150.50	Matches:	64
Percent Similarity:	36.27%	Conservative:	10
Best Local Similarity:	31.37%	Mismatches:	53
Query Match:	12.94%	Indels:	77
DB:	2	Gaps:	11

US-10-030-225-1\_COPY\_58\_696 (1-639) x T33110 (1-460)

Qy 3 GCGTGCCTCCGCGA-----CCGAGCGGCTCTCTCTGCTGCTGCTGCGGCGC 56

Db 298 AlaAlaAlaProGluAlaAlaProGluAlaAlaProGluAlaAlaGluGlyAlaGlyGly 317

Qy 57 GCTGCTGGAGCGCGCTAGGGCTCCCGTGAAGACCGCGGCTCCGCGACCGGCC 116





Db 1926 GlyGlyTyrgly 1929  
||| |||  
RESULT 6  
H75457  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C/Species: Deinococcus radiodurans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: H75457  
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A/Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.  
A/Reference number: A75250; PMID:20036896; PMID:10567266  
A/Accession: H75457  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-395 <WHI>  
A/Cross-references: UNIPROT:Q9RVT4; GB:AE001946; GB:AE000513; NID:G6458655; PIDN:AAFI051  
A/Experimental source: strain R1

```

C:Genetics:
A:Gene: DR0938
A:Map position: 1

Alignment Scores:
Pred. No.: 0.0853 Length: 395
Score: 128.00 Matches: 64
Percent Similarity: 37.26% Conservative: 15
Best Local Similarity: 30.19% Mismatches: 80
Query Match: 11.01% Indels: 53
DB: 2 Gaps: 10

US-10-030-225-1_COPY58_696 (1-639) x H75457 (1-395)

QY 18 CCGAGCCGGGCTCTCTCTCTGCTGCTGCTGCT----- 50
Db 23 ProArgArgTrpProGluLeuValAlaAlaThrLeuValLeuAlaLeuGlyValGlyLeu 42
QY 51 GCGGGCGTCTGGAGGGCGGCTAGGGTCCCGGTGAAGAGCCGGGCTCCGGGACC 110
Db 43 GlySerLeuLeuGlyGluArgAlaProAlaGluProValGlyAlaProGlnPro 62
QY 111 ACGGCTGGGAGCTCACGAGGCTCCGAGGCTCTCAGGCTCCCGAGATCCTAGGCTCT 170
Db 63 ThrThr-----SerileProValSerProGlyAla 72
QY 171 GAAGGAAGAGGAGGAGGC-----ACCATGCTGCC 200
Db 73 ValValAlaAspGlySerProGlnThrAlaAlaProAlaGlnAlaGlnThrAlaGly 92
QY 201 CAGAACCCACTGCAGGAGAGCCACCAACATGATGCTGGACT-----GTCAC 251
Db 93 LysileProProAlaProAlaProAlaProLysile-ProProProValLeuPr 112
QY 252 TGAGCCAGCAGCATGACCCAGGCAACACCCCTCCAGAGACCCAGAGTTACT-- 309
Db 112 oGluProArg-gme----ProAlaProThrProProArgProValGluGluThrTh 131
QY 310 -----CCGTGTCGAGTGTGAGAAAGCTGCCGGGATTGGCCAGCAC 353
Db 131 rThrAlaAlaSerProIleThrGlnProProValThrGlnAlaProAlaThrGlnThrPr 151
QY 354 AACCTTGAGTACCCCT-----AACCTGANTACCCAGGTTTCAGCTCCCGCATCC 404
Db 151 oThrProGlnThrProAlaAlaGlnAlaProAlaThrGlnProProAlaThrProAlaPr 171
QY 405 TAGGCTCTGAGGAAGAGGAGGACGACTGTCTCCCGAGAACCCACCTGCGAGCAGA 464
Db 171 oGluProAlaAlaProAlaGluProAla-----ProAlaThrThrProGluProAl 188
QY 465 GCTACACCAACATGGATGTTGGCTTCACTGAGCCAGCAGCCCTGACCCCGGGGAATGC 524

```

EDBEXD  
immediate-early protein RL2 - human herpesvirus 2 (strain HG52)  
N;Alternate names: RL2 protein  
C;Species: human herpesvirus 2  
A;Note: host Homo sapiens (man)  
C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
C;Accession: JQ1501  
R;McGeoch, D.J.; Cunningham, C.; McIntyre, G.; Dolan, A.  
J. Gen. Virol. 72, 3057-3075, 1991  
A;Title: Comparative sequence analysis of the long repeat regions and adjoining parts of  
F;122-172/Domain: RING finger homology <RG>  
F;126-166/Region: zinc finger C3HC4 motif  
A;Reference number: JQ1494; MUID:92113549; PMID:1662697  
A;Accession: JQ1501  
A;Molecule type: DNA  
A;Residues: 1-825 <MCG>  
A;Cross-references: UNIPROT:P28284; GB:D10471; DDBJ:D01128; NID:g221784; PIDN:BAA23427.1  
C;Genetics:  
A;Gene: RL2  
A;Introns: 25/3; 252/1  
C;Superfamily: herpesvirus immediate-early protein IF110; RING finger homology  
C;Keywords: DNA binding; immediate-early protein; tandem repeat; transcription regulation  
F;122-172/Domain: RING finger homology <RG>  
F;126-166/Region: zinc finger C3HC4 motif  
F;589-623/Region: 5-residue repeats (A-S-S-S-S)

Alignment Scores:  
Pred. No.: 0.104 Length: 825  
Score: 126.50 Matches: 65  
Percent Similarity: 34.84% Conservative: 20  
Best Local Similarity: 26.64% Mismatches: 92  
Query Match: 10.88% Indels: 67  
DB: 1 Gaps: 13

US-10-030-225-1\_COPY\_58\_696 (1-639) x EDBEXD (1-825)

Qy	3	CGCTGCGCTCGCGACGCGCGGCTCCT	-----CCTCTG 38
Db	281	AlaLathArgProAlaProGlyAlaProArgSerSerSerGlyGlyAlaProLeu 300	
Qy	39	CGTCTGCTGCTGGCGGCGCTGCTGAGCGCGCTAGGCTCCCGTGAAGAAGCCGCG 98	
Db	301	ArgAlaGlyValGlySerGlySerGlyGlyGly-----ProAlaValAlaAlaVal 317	
Qy	99	GCTCCGCGGACACGCGCTCGGAGGCTCACGAGGCTCGCAGAGTCTCAGCCCTCCCGAGA 158	
Db	318	ValProArgValAlaSerLeuProPro---AlaAlaGlyGlyArgAlaGlnAlaArg 336	
Qy	159	-----TCCTAGGCTCTGAGAAGAGGAGGAGGACCACTGCTCCCGCAG 203	
Db	337	ArgValGlyGluAspAlaAlaAlaGluGlyArg-----ThrProAla 352	
Qy	204	AACCCACTGCGAGGACGACACCAATGATGCTGACTGTCTCAGCCAGCCAGCAGC 263	
Db	353	ArgGlnProArgAlaAlaGlnGluProProLleValle-----SerAspSer 368	
Qy	264	CATGACCCCGAGCAACACCCCTCC---CAGGACCCAGAGGTTACTCCGTTCGCGCT 320	
Db	369	ProProSerProArgArgProAlaGlyProGlyProLeuSerPheValSerSerSer 388	
Qy	321	GGAGCTGCAAGAGTGC CGGGATTGGCCAGCACACCTTGAGTACCCCTAACCTGTATC 380	
Db	389	SerAlaGlnValSerSerGlyProGly----- 397	
Qy	381	CCAGGCTTCAGCTCCCGATCCTAGGCTCTGAGGAAGAGGAGGAGGACGACGCTGCT 440	
Db	398	---GlyGlyGlyLeuProGlnSerSerGlyArgAlaAlaArgProArgAlaAlaValAla 416	
Qy	441	CCC-----CAGAACCCACCT----- 455	
Db	417	ProArgValArgSerProProArgAlaAlaAlaProValValSerAlaSerAlaAsp 436	
Qy	456	---GCAGCGAGAGTACACCAATGGATGTTGGACTGTCTCAGTGCACGACCCCTGAC 512	
Db	437	AlaAlaGlyProAlaProAlaValProValAspAlaHis-----ArgAlaProArg 454	





Db	1682	-----AlaGluGluProGlnProAlaIuYbSseGluAlaAlaProProAlaAsnGlnPr	1699
Qy	196	-----CTCCCCAGAACCCACCTCTGAGGAGCAGCCACACCACTGGATCGTGAC	245
Db	1699	oPheSerArgPheGlnValThrProIleGlnAlaSerProValLeuHis	1715
Qy	246	TGTCACTAGCCAGCAGCCATGATCCCGAGCAACACCACTCCAGGACCCAGAGGT	305
Db	1716	-----ThrLeuValGlnArgSerProThrVa	1724
Qy	306	TACTCTCGTTCGGCTGGAGCTGCAGAACTGCCGGATTGGCCAGCACA	354
Db	1724	lMetProAlaProLeuProProAlaProValAlaAlaSerAlaSerAlaAlaProSerPr	1744
Qy	355	-----ACCTTGAGTACCCCTTAACCTGATACCCAGGCTTCAGCTCCCCAGATCCTAGGCC	410
Db	1744	oGlnThrProSerArgProGlyProAlaValSerGlnThrMetSerProValProHisPr	1764
Qy	411	TCCTGAGGGAA-----GAGGAGGAGCGACTGCTGCCCCAG	446
Db	1764	oLeuArgGlnProThrAlaThrPheValPheThrGluArgGluGlyGluProIleProVa	1784
Qy	447	AAOCCCACTGCAGCGCAGAGCTACACCAACATGGAGTGTGGACTGTCACTGAGCCAGCAGC	506
Db	1784	lSerGlnProProAlaGlnThr--GlnHisGlnProValArgIleSerGlnIlysthrAl	1803
Qy	507	CTGACCCCGAGGAATGCCAGGCT-----	531
Db	1803	aProValProSerSerSerMetProSerAlaSerGluAlaMetProArgSerAlaGlyTr	1823
Qy	532	-----CCAGGACCCAGGAGGTACTCCCTTGCTGCTGGAGCTGCAGAAGCTGCCAGAA	587
Db	1823	pIleProGluAlaAsnArgProThrProLeuLeuSerGlnGlnHisGluLeuArgGluVa	1843
Qy	588	G 588	
Db	1843	i 1843	
RESULT 12			
Q088			
BPLF1 protein - human herpesvirus 4 (strain B95-8)			
C/Species: human herpesvirus 4, Epstein-Barr virus			
C/Date: 25-Feb-1985 #sequence revision 25-Feb-1985 #text_change 09-Jul-2004			
C/Accession: G93065; A03747; S32993			
R/Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.			
Mol. Biol. Med. 1, 21-45, 1983			
A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-B			
A/Reference number: A93065; MUID:85035713; PMID:6092825			
A/Accession: G93065			
A/Molecule type: DNA			
A/Residues: 1-3149 <BAN>			
A/Cross-references: UNIPROT:P03186; EMBL:V01555; NID:G59074; PIDN:CAA24839.1; P			
R/Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson			
Nature 310, 207-211, 1984			
A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.			
A/Reference number: A03794; MUID:84270667; PMID:6087149			
A/Contents: annotation; protein coding region			
C/Superfamily: human herpesvirus 4 BPLF1 protein			

```

Alignment Scores:
Pred. No.:          0.12          Length:          3149
Score:             125.00         Matches:          65
Percent Similarity: 39.57%        Conservative:    28
Best Local Similarity: 27.66%     Mismatches:     85
Query Match:       10.75%        Indels:         57
DB:                1            Gaps:            15

US-10-030-225-1_COPY58_696 (1-639) x QQBEB (1-3149)

Qy      12  CCSCGACCGAGCCGGGCTCTCTCTGCTGCTGCTGCGCGCGCTGCTGAGGGGC 71
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      395  ProAlaProSerThrProPro-----ArcAlaSerSerGlyValala----- 408

```



J. Biol. Chem. 268, 2836-2843, 1993  
 A:Title: Molecular characterization of the family of the N-methyl-D-aspartate receptor  
 A:Reference number: A45219; MUID:93155102; PMID:8428958  
 A:Accession: D45219  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1265-1323 <ISH>  
 A:Experimental source: brain  
 A:Note: sequence extracted from NCBI backbone (NCBIP:124265)  
 C:Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology  
 F:451-879/Domain: glutamate receptor homology <GRH>

Alignment Scores:  
 Pred. No.: 0.165 Length: 1323  
 Score: 123.50 Matches: 74  
 Percent Similarity: 35.27% Conservative: 17  
 Best Local Similarity: 28.68% Mismatches: 85  
 Query Match: 10.62% Indels: 83  
 DB: 2 Gaps: 15

US-10-030-225-1\_COPY\_58\_696 (1-639) x I78557 (1-1323)

```

Qy 9 GCTCGGACGAGCGGGCTCTCT-----CCTCTGCGTGTGCTGCTGGCGGC 56
Db 899 AlaProProAlaAlaArgProProProProGlnProLeuProSerProAlaIlePro 918
Qy 57 GCTGCT-----GGAGGCGGCTAGGCTCCCGCTGAAGAAGCGCGGCTCCG 104
Db 919 AlaAlaArgProProGlyProAlaProPheValProArgGluArgAlaAlaAsp 938
Qy 105 CGGACCGGCTGG-----TTPArgAlaAlaLysGlyThrGlyProProGlyGlyAlaAlaIleAla 955
Db 939 Arg-----TTPArgAlaAlaLysGlyThrGlyProProGlyGlyAlaAlaIleAla 955
Qy 120 -----GAGCTCAGAGGCTCGCAGAGGCTCTCAGC 149
Db 956 AspGlyPheHisArgTyrTyrGlyProIleGluPro-----GlnGlyLeuGly 971
Qy 150 CTCGCCAGATCTAGGCTCTGAAGAAGAGGAGGAGC-----188
Db 972 LeuGlyGluAlaArgAlaAlaProArgGlyAlaAlaGlyArgProLeuSerProProThr 991
Qy 189 ACCACTGCTCCAGAACCCACTGCGAGGAGAGCCACACCAATGATGCTGGACTGT 248
Db 992 ThrGlnProProGlnLysProPro-----ProSerTyrPheAla-IleVa 1006
Qy 249 CACTGAG-----CCAGCAGCCATGACCCAGGAGCTCGAAGAGTCCGGGATGGCCAG 290
Db 1006 ArgGluGlnGluProThrGluProProAlaAlaGlyAlaPheProGlyPheProSerProPr 1026
Qy 291 CAGGACCCAGAGGTTACTCCGCTTGGAGCTGCGAGGCTCGAAGAGTCCGGGATGGCCAG 350
Db 1026 AlaProProAlaAlaAlaAlaAlaValGly-----ProProLeuGlyArgLeuAlaPh 1045
Qy 351 CACAACCTTGATACCCCTAACCTGTATACCCAGGCTTCAGCCTCCCGAGATCTAGGCC 410
Db 1045 eGluAspGluSerProProAlaProSerArgTyrProArgSerAspProGluSerGlnPr 1065
Qy 411 TCTGAG-----GGAAGAGGAGGAGGAGGAGCAGCAGT-----GCTCCCGAGAACCCACCTGCAGG 460
Db 1065 oLeuLeuGlyGlyAlaGlyProSerAlaGlyAlaProThrAlaProProProAr 1085
Qy 461 CAGAGCTACA-----CCAACATGGATGTGGAGTGTCACTGAGCAGCAGCCCTGACCC 514
Db 1085 gArgAlaAlaProProProCysAlaTyrLeuAspLeu-----GluProSerPr 1101
Qy 515 CAGGAATGCCAGCGCTCCAGGAGCCAGGAGGTTACTCTTCTGCTGAGTGCAGAG 574
Db 1101 oSerAspSerGluAspSerGluSerLeuGlyAlaSerLeu-----GlyGlyLeuG 1119
Qy 575 A-----GCTCCCGAGAAATGGTCCA 593
Db 1119 uProTrpTrpPheAlaAspPheProTyrProTyrAlaGluArgLeuGlyPro 1136

```

# RESULT 15

C45219  
 N-methyl-D-aspartate receptor chain NMDAR2D-1 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: C45219  
 R:ishi, T.; MoriYoshi, K.; Sugihara, H.; Sakurada, K.; KadoTani, H.; Yokoi, M.; Akazawa  
 J. Biol. Chem. 268, 2836-2843, 1993  
 A:Title: Molecular characterization of the family of the N-methyl-D-aspartate receptor  
 A:Reference number: A45219; MUID:93155102; PMID:8428958  
 A:Accession: C45219  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-1356 <ISH>  
 A:Experimental source: brain  
 A:Note: sequence extracted from NCBI backbone (NCBIP:124264)  
 C:Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology  
 F:451-879/Domain: glutamate receptor homology <GRH>

Alignment Scores:  
 Pred. No.: 0.164 Length: 1356  
 Score: 123.50 Matches: 74  
 Percent Similarity: 35.27% Conservative: 17  
 Best Local Similarity: 28.68% Mismatches: 85  
 Query Match: 10.62% Indels: 83  
 DB: 1 Gaps: 15

US-10-030-225-1\_COPY\_58\_696 (1-639) x C45219 (1-1356)

```

Qy 9 GCTCGGACGAGCGGGCTCTCT-----CCTCTGCGTGTGCTGCTGGCGGC 56
Db 899 AlaProProAlaAlaLysProProProProGlnProLeuProSerProAlaIlePro 918
Qy 57 GCTGCT-----GGAGGCGGCTAGGCTCCCGCTGAAGAAGCGCGGCTCCG 104
Db 919 AlaAlaArgProProGlyProAlaProPheValProArgGluArgAlaAlaAsp 938
Qy 105 CGGACCGGCTGG-----TTPArgAlaAlaLysGlyThrGlyProProGlyGlyAlaAlaIleAla 955
Db 939 Arg-----TTPArgAlaAlaLysGlyThrGlyProProGlyGlyAlaAlaIleAla 955
Qy 120 -----GAGCTCAGAGGCTCGCAGAGGCTCTCAGC 149
Db 956 AspGlyPheHisArgTyrTyrGlyProIleGluPro-----GlnGlyLeuGly 971
Qy 150 CTCGCCAGATCTAGGCTCTGAAGAAGAGGAGGAGC-----188
Db 972 LeuGlyGluAlaArgAlaAlaProArgGlyAlaAlaGlyArgProLeuSerProProThr 991
Qy 189 ACCACTGCTCCAGAACCCACTGCGAGGAGAGCCACACCAATGATGCTGGACTGT 248
Db 992 ThrGlnProProGlnLysProPro-----ProSerTyrPheAla-IleVa 1006
Qy 249 CACTGAG-----CCAGCAGCCATGACCCAGGAGCTCGAAGAGTCCGGGATGGCCAG 290
Db 1006 ArgGluGlnGluProThrGluProProAlaAlaGlyAlaPheProGlyPheProSerProPr 1026
Qy 291 CAGGACCCAGAGGTTACTCCGCTTGGAGCTGCGAGGCTCGAAGAGTCCGGGATGGCCAG 350
Db 1026 AlaProProAlaAlaAlaAlaAlaValGly-----ProProLeuGlyArgLeuAlaPh 1045
Qy 351 CACAACCTTGATACCCCTAACCTGTATACCCAGGCTTCAGCCTCCCGAGATCTAGGCC 410
Db 1045 eGluAspGluSerProProAlaProSerArgTyrProArgSerAspProGluSerGlnPr 1065
Qy 411 TCTGAG-----GGAAGAGGAGGAGGAGGAGCAGCAGT-----GCTCCCGAGAACCCACCTGCAGG 460
Db 1065 oLeuLeuGlyGlyAlaGlyProSerAlaGlyAlaProThrAlaProProProAr 1085
Qy 461 CAGAGCTACA-----CCAACATGGATGTGGAGTGTCACTGAGCAGCAGCCCTGACCC 514
Db 1085 gArgAlaAlaProProProCysAlaTyrLeuAspLeu-----GluProSerPr 1101

```

```

Qy      515 CAGGGAATGCCACGCTCCACGAGCCACGAGAGTTACTCCCTTCTGCTGGAGCTGCAGA 574
Db      1101 oSerAspSerGluAspSerGluSerLeuGlyGlyAlaSerLeu-----GlyGlyLeuGl 1119
Qy      575 A-----GCTGCCAGAAATGGTCCA 593
Db      1119 uProTyrTrpPheAlaAspPheProTyrProTyrAlaGluArgLeuGlyPro 1136

```

Search completed: February 19, 2005, 00:23:05  
Job time : 51 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 20:52:54 ; Search time 426 Seconds  
(without alignments)  
8879.612 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 639  
Sequence: 1 atgcgtgcgtccgcagcgc.....ataaccaggtagaccatcaag 639

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004as:\*
- 13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639	100.0	2981	4	Aaf29348 Human pro
2	639	100.0	2981	5	Aaf93820 Human cdn
3	542.6	84.9	750	5	Aaf94019 Primer sp
C 4	261.8	41.0	497	4	Aai10232 Probe #16
C 5	261.8	41.0	497	4	Aba51868 Human foe
C 6	261.8	41.0	497	4	Aai31480 Probe #16
C 7	261.8	41.0	497	4	Aba21690 Probe #15
C 8	261.8	41.0	497	4	Aak25606 Human bon
C 9	261.8	41.0	497	4	Aak00167 Human bra
C 10	261.8	41.0	497	4	Abs25182 Human liv
C 11	261.8	41.0	497	5	Aai00170 Probe #16
C 12	261.8	41.0	497	6	Abs00175 Human gen
C 13	239.8	37.5	243	4	Aai19514 Probe #94
C 14	239.8	37.5	243	4	Aba64531 Human foe
C 15	239.8	37.5	243	4	Aai44705 Probe #13
C 16	239.8	37.5	243	4	Aba46659 Human bre
C 17	239.8	37.5	243	4	Aba31664 Probe #10
C 18	239.8	37.5	243	4	Aak38709 Human bon
C 19	239.8	37.5	243	4	Aak12982 Human bra
C 20	239.8	37.5	243	4	Abs38283 Human liv

C 21	239.8	37.5	243	5	Aai05236	Aai05236 Probe #52
C 22	239.8	37.5	243	6	ABSI2780	ABSI2780 Human gen
C 23	141	22.1	522	10	ADFS8409	ADFS8409 Human pol
C 24	91.8	14.4	1128	6	ABQ30568	ABQ30568 Oligonuc1
C 25	91.8	14.4	1128	6	ABQ30569	ABQ30569 Oligonuc1
C 26	85.6	13.4	1128	6	ABQ30566	ABQ30566 Oligonuc1
C 27	85.6	13.4	1128	6	ABQ30567	ABQ30567 Oligonuc1
C 28	44.8	7.0	1047	11	ABDI0430	ABDI0430 Pseudomon
C 29	44.8	7.0	1464	11	ABDI0367	ABDI0367 Pseudomon
C 30	44.8	7.0	1494	11	ABDI0513	ABDI0513 Pseudomon
C 31	44.8	7.0	2704	12	ADQ64839	Adq64839 Novel hum
C 32	42.6	6.7	3018	4	AAH78274	AAH78274 Coding se
C 33	42.6	6.7	3391	4	AAH78273	AAH78273 Nucleotid
C 34	42.6	6.7	5968	4	AAS27841	AAS27841 DNA encod
C 35	42.6	6.7	5968	4	AAS35091	AAS35091 DNA #41 e
C 36	42.6	6.7	5968	4	AAK80295	AAK80295 Human imm
C 37	42.6	6.7	5968	10	ADB94644	ADB94644 Novel hum
C 38	42.6	6.7	5968	10	ADC46533	ADC46533 Human neo
C 39	42.2	6.6	815	3	AAA96721	AAA96721 Polynucle
C 40	42.2	6.6	3334	13	ADR08320	ADR08320 Full leng
C 41	41.8	6.5	15872	2	AAT68715	AAT68715 Streptomy
C 42	41.6	6.5	4705	8	ACA61566	ACA61566 Streptomy
C 43	41.6	6.5	4736	8	ACA61565	ACA61565 Streptomy
C 44	41.6	6.5	8174	6	AAL44951	AAL44951 Equine he
C 45	41.6	6.5	149158	12	ADP74211	ADP74211 Equine he

#### ALIGNMENTS

#### RESULT 1

AAF29348  
ID AAF29348 standard; cdna; 2981 BP.

XX AAF29348;

XX 20-APR-2001 (first entry)

XX Human proliferation differentiation factor cdna sequence.

XX Human; proliferation differentiation factor; haematopoietic function; ss.

XX Homo sapiens.

XX WO200104312-A1.

XX 18-JAN-2001.

XX 06-JUL-2000; 2000WO-JP004514.

XX 08-JUL-1999; 99JP-00194179.

XX 18-OCT-1999; 99US-0159586P.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;

XX WPI: 2001-138354/14.

XX P-PSDB; AAB49765.

XX Proliferation-differentiation factor protein encoded by PSEC137 cloned from human cdna library, being hematopoietic factor for inducing differentiation of blood cells, used to maintain hematopoietic function.

XX Claim 1; Page 36-41; 49pp; Japanese.

XX This invention relates to polynucleotide sequence PSEC137, which encodes a human proliferation differentiation factor protein. Included in the invention is a vector containing the cdna sequence, a transformant containing the vector, and a process for producing the protein. The protein together with its encoded DNA and antisense DNA are used in drugs to maintain the haematopoietic function. The present sequence represents cdna encoding the proliferation differentiation factor protein

















RESULT 12  
ABS00175/c  
ID ABS00175 standard; DNA; 497 BP.  
XX  
AC ABS00175;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human genome-derived single exon probe from lung SEQ ID No 166.  
XX  
DE Human; ds; single exon probe; asthma; lung cancer; COPD; ILB;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000665.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PI WPI; 2002-114183/15.  
XX  
PT Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
XX  
PS Claim 1; SEQ ID NO 166; 634pp; English.  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human





**This Page Blank (uspto)**



Result No.	Score	Query			ID	Description
		Match	Length	DB		
C 1	44.8	7.0	1047	4	US-09-252-991A-9034	Sequence 9034, Ap
	44.8	7.0	1464	4	US-09-252-991A-8971	Sequence 8971, Ap
	44.8	7.0	1494	4	US-09-252-991A-9117	Sequence 9117, Ap
C 5	43.6	6.8	1494	4	US-09-302-540-6461	Sequence 6461, Ap
	43.6	6.8	1664	4	US-09-502-540-452	Sequence 452, App
	42.2	6.6	815	3	US-09-383-586-9	Sequence 9, Appli
6	42.2	6.6	815	4	US-09-823-038A-9	Sequence 9, Appli
8	41.2	6.4	732	4	US-09-252-991A-8860	Sequence 8860, Ap
9	41	6.4	2283	3	US-09-252-991A-3298	Sequence 3298, Ap
10	39.8	6.2	15872	3	US-09-105-537-1	Sequence 1, Appli
11	39.8	6.2	15872	4	US-09-091-609-3	Sequence 1, Appli
12	39.8	6.2	15872	4	US-09-091-609-3	Sequence 3, Appli
13	39.8	6.2	43280	2	US-08-804-227C-1	Sequence 1, Appli
C 14	39.4	6.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	38.8	6.1	40586	4	US-09-949-016-16965	Sequence 16965, A
	38.6	6.0	1638	4	US-10-237-551-119	Sequence 119, App
16	38.6	6.0	1638	4	US-10-237-551-148	Sequence 148, App
17	38.6	6.0	1644	4	US-10-237-551-213	Sequence 213, App
18	38.6	6.0	1644	4	US-10-237-551-214	Sequence 214, App
19	38.6	6.0	1644	4	US-09-827-688-8	Sequence 8, Appli
20	38.6	6.0	154746	4	US-09-252-991A-6380	Sequence 6380, Ap
21	38.4	6.0	621	4	US-09-252-991A-6169	Sequence 6169, Ap
22	38.4	6.0	933	4	US-09-252-991A-6012	Sequence 6012, Ap
C 23	38.4	6.0	1350	4	US-09-252-991A-6012	Sequence 6012, Ap
24	38.2	6.0	1393	2	US-08-765-875-1	Sequence 1, Appli
25	38.2	6.0	1393	3	US-08-795-671-1	Sequence 1, Appli
26	38.2	6.0	1393	4	US-09-454-540-1	Sequence 1, Appli
27	38.2	6.0	1393	4	US-09-626-896-24	Sequence 24, Appl

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```

, FILE REFERENCE: 107196.136
, CURRENT APPLICATION NUMBER: US/09/252,991A
, CURRENT FILING DATE: 1999-02-18
, PRIOR APPLICATION NUMBER: US 60/074,788
, PRIOR FILING DATE: 1998-02-18
, PRIOR APPLICATION NUMBER: US 60/094,190
, PRIOR FILING DATE: 1998-07-27
, NUMBER OF SEQ ID NOS: 33142
, SEQ ID NO 8971
, LENGTH: 1464
, TYPE: DNA
, ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8971

```

	Query Match	7.0%;	Score 44.8;	DB 4;	Length 1464;
	Best Local Similarity	50.0%;	Pred. No. 0.11;		
	Matches 112;	Conservative 0;	Mismatches 112;	Indels 0;	Gaps 0;
QY	7	GCGCTCCGACCGAGCGGGCTCTCTCTGTCGTCTGCTGTGCGGGCGCTCTCTGGAG	66		
DB	685	GAGCGCGCAAGCTGGCGACGGATTTCACCGCGCTGCTCTGGTTGCCGCTCTCGGC	744		
QY	67	GCGCGCTGAGGCTTCCCGTGAAAGAGCGCGGGTCCGCGGACACCGCTCTGGAGGCTC	126		
DB	745	CAGCGCAGGCGCGCGCGAGGAACAGCGCTTCTCCGCGCCCGCGCAAGCGACTG	804		
QY	127	ACGAGGCTCCAGAGGTCCTCAGCTCTCCAGATCCTTAGGCTCTGAGGAGAGAGGAGGAG	186		
DB	805	CACGCTGTCCACAGGGCGACTCCGCGAGGATCAGTGTGCTCGCGGACGTAGGGCAG	864		
QY	187	GCACCACTGTCTCCCAAGACCCACTCTCAGGCAGAGCCACACCA	230		
DB	865	GCTGACGTAGCGCGCGCTCTCTTCGCGCAGCGCAGCAACGGCA	908		

## RESULT 3

```

US-09-252-991A-9117/c
; Sequence 9117, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9117
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9117

```

Query Match	7.0%;	Score 44.8;	DB 4;	Length 1494;
Best Local Similarity	50.0%;	Pred. No. 0.11;		
Matches 112; Conservative	0;	Mismatches 112;	Indels 0;	Gaps 0;
QY	7	GC	CTCGGCACGAGCGGGCTCCTCCTCTGCGTGTCTGTCTGCGGGCGCTGCTGGAG	66
Db	879	GAGCGGCGAAGCTGGCGACGGAATTTCCACGGGTCTGTCTGTGTTGCCGCTCTCGCG	820	
QY	67	GC	GGCGGTAGGGGTTCCCGCTGAAGAGCCGGCTCCGGCGACACACGGCTTGGAGGCTC	126
Db	819	CAGCGCAGCGCCCGCGAGGACAGCGCTTCTCCGGGCCCGCGCAAGCACTG	760	
QY	127	AC	GAGGCTCGCAGGTTCTCAGCCTCCCAAGATCTTAGGCTCTGAAGGAGGAGGAG	186
Db	759	CAGCGCTCTCCCAGGGCGCACTCCGGCAGGATCAGTGGCGTCCGGCGGACGTAGGCGAG	700	

Qy 187 GCACCACTGTTCCCTCCAGAACCCACCTGCAGGACAGCCACACCA 230  
Dβ 699 GCTGACGTAGGCGCCGCCCTCTTCGCCACGGCGAGCAACGGCA 656

## RESULT 4

```

US-09-902-540-6461
; Sequence 6461, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6461
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/key: unsure
; LOCATION: (1)..(1494)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-6461

```

Query Match 6.8%; Score 43.6; DB 4; Length 1494;

Query match: 100%; Score 43.07; DB 4; Length 145;  
Best Local Similarity 46.3%; Pred. No. 0.22;  
Matches 215; Conservative 0; Mismatches 244; Indels 5; Gaps 2;  
QY 16 GACCGAGCCGGCTCTCTCTCGGCTCTCTGCTGCGCGGCTGCTGAGGCGCGCTA 75  
Db 550 GACAAGCGAGCTCATCCACTCATGCTCCAGTGGCGGCTGAGAGCCCGCTGGT 609  
QY 76 GGCGTCCCGTGAAGAGCCGCGGCTCCGCGGACCACGGCGCTGGAGACCTCACGAGGTC 135  
Db 610 CTGCTGGCATCGCAACTTGGGCTCTATTGGCGAGAGCCCGCGAGCCCTCGCGGCTTC 669

QY	136	GCAGAGTCTCAGCCTCCCGAGCTCTGAAGCAGGAGGAGGACCACTG	193
Db	670	AACCA CGCATCGCTTAC --- GTGCCCAAGTTGACCTGTACCTGGACGCACCGCGGAG	726
QY	196	CTTCCAGAACCACTCTGAGGAGAGCACAACAATGATGCTGGACTGTCTACTGAG	255
Db	727	TTCCACGGGCGCCAACAGAGTGCACGCGCGAGCCGGGTGGCCAACTGCTGGTGGTGGAG	786
QY	256	CCAGCAGCATGACCCAGGCAACACACCCCTCCAGGACCCACAGAGGTTACTCCGTTG	315
Db	787	CCCAACGGCAGAGCACCTTCTCAACACGCGGAGGCGAAGGGCGGAGCACGCCACG	846
QY	316	CGGCTGGAGTGCAGAAGTGC CGGGATTGGCCAGCAACAACCTTGAGTACC -- CCTAACCC	373
Db	847	CGCATCTCGTGGACGTGACGCTCCGCGCGGAGCGGCGGCGCTGAAATCACCGGCTCCAGC	906
QY	374	CTGATACCCAGGCTTCAGCCTCCCCAGATCCTTAGGCCTCTGAGGAAGAGGAGGAGGCAC	433
Db	907	TCCGTGGGCGGAGCAGCGGCGCCGAGCTACCGCGCGCCTTACCGCGCGAGGCCACGCGC	966
QY	434	GACTGCTCCCCAGAACCCACCTCGAGCGAGACTACACCAACAT	477
Db	967	AAGTCCACCTTCAGCGCGCGCTGGGCGCAGAGCTTCCCGGCGCT	1010

## RESULT 5

US-09-902-540-452/c  
; Sequence 452, Application US/09902540  
; Patent No. 6833447

```

; CURRENT APPLICATION NUMBER: US/09/383,586
; CURRENT FILING DATE: 1999-08-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 815
; TYPE: DNA
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (119)...(682)
; NAME/KEY: sig_peptide
; LOCATION: (119)...(205)
US-09-383-586-9

Query Match
Best Local Similarity 6.6%; Score 42.2; DB 3; Length 815;
Matches 56; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 7 GCGCTCCGCGACCGAGCGGGCTCTCTCTGCTGCTGCTGCTGCGGGCGCTCTGAG 66
Db 122 GCGCTGTGCGCGCGCGCGCGCTCTCTCTGCTGCGGGCTGCTGCGAGTGTCTGCTAGG 181

QY 67 GCGGCGCTAGGGCTCCCCG 85
Db 182 GCGGCGCAGGACCGACCG 200

RESULT 7
US-09-823-038A-9
; Sequence 9, Application US/09823038A
; Patent No. 6797271
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Abernethy, Nevin
; APPLICANT: Onrust, Rene
; APPLICANT: Kumble, Anand
; APPLICANT: Murison, Greg
; TITLE OF INVENTION: Compositions Isolated From Stromal Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1037c3
; CURRENT APPLICATION NUMBER: US/09/823,038A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 815
; TYPE: DNA
; ORGANISM: Mouse
US-09-823-038A-9

Query Match
Best Local Similarity 6.6%; Score 42.2; DB 4; Length 815;
Matches 56; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 7 GCGCTCCGCGACCGAGCGGGCTCTCTCTGCTGCTGCTGCTGCGGGCGCTCTGAG 66
Db 122 GCGCTGTGCGCGCGCGCGCTCTCTCTGCTGCGGGCTGCTGCGAGTGTCTGCTAGG 181

QY 67 GCGGCGCTAGGGCTCCCCG 85
Db 182 GCGGCGCAGGACCGACCG 200

RESULT 8
US-09-252-991A-8860
; Sequence 8860, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```



[illegible][illegible]

Db 36635 CTTCTGCTGAGGTCTCTGGAGGCGCTGGAACGCGCGGCATCGACCGCTCTCGCT 36694  
Qy 102 CCGGGACACCGGCTGGAGCCTCACAGGCTCGCAGAGTCTCAGCCTCCCGATCC 161  
Db 36695 GCGGGACGCGCGTGTGTCTACGTGGCGCGCGCACGGCTGTACGCTCCGATCC 36754  
Qy 162 TAGGCTCTGAAGGAAGAGGAGGACCACTGCTCCCGAAGACCCACCTGCGAGGAGA 221  
Db 36755 CCGCTGTGCTCCGAGGCTCGAGGGCTATCTGTCACCGCAGCGCGCGGTGAT 36814  
Qy 222 GCCACACCAATGATCTGATGTCTACGTAGCAGCAGCAGCAGCAGCAGCAGCAGC 281  
Db 36815 GTCCGGCGCATCTCTACGGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGC 36874  
Qy 282 CACCCTCCAGGACCCAGAGGTACTCCGTTGGGCTGGAGTGCAGAA 332  
Db 36875 CTGCTCTCTCGTGTGGGCTGCATCTGGCGGTACGGGCGCTGGCGCA 36925

## RESULT 14

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 6.2%; Score 39.4; DB 3; Length 4403765;  
Best Local Similarity 52.8%; Pred. No. 19;  
Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
Qy 28 CTCTCTCTCTGCTGCTGCTGCTGCGCGCTGCTGAGGCGCGCTAGGCTCCCGTG 87  
Db 285911 CCCCTGCCAAGTACACAAAGTGGGCGGTGATCCGCTGCCCTGCTGGGCTTG 285852  
Qy 88 AAGAGCCGCGGCTCCGCGACCAAGGCTCGGAGCCTCACAGGCTCGCAGAGGTCTCA 147  
Db 285851 GCGAGCTGCTGAGCGGAGTTCCGCTGCCAGCAGCGCGCGCGCTGGCTGCGC 285792  
Qy 148 GCCTCCCGAGATCTAGGCTCTGAGGAAGAGGAGGAGGC 188  
Db 285791 GCGTTCGCGCACCCCGAGCGGACAAAGCAGTAGCGGTGGC 285751

## RESULT 15

US-09-949-016-16965  
; Sequence 16965, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01307  
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16965  
; LENGTH: 40586  
; TYPE: DNA  
; ORGANISM: Human  
; US-09-949-016-16965

Query Match 6.1%; Score 38.8; DB 4; Length 40586;  
Best Local Similarity 51.8%; Pred. No. 8.2;  
Matches 88; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
Qy 50 TGGCGGCGCTGCTGAGGCGCGCTAGGGCTCCCGTGAAGAAGCCGCGGCTCCCGGAC 109  
Db 2143 TGGCGGCGCTGCCGAGACGAGCTCAGTTCCCGAGGAGGCTCGGAGCTCGGGCAG 2202  
Qy 110 CACGCGCTGGAGCCTCACAGGCTTCGAGAGTCTCAGCCTCCCGAGATCCTAGGCTC 169  
Db 2203 AGCCTGCGCGCGCGCTCGAGGCGCGGATGAGCCTCCCGCTCTGCTTGGCTTG 2262  
Qy 170 TGAAGGAAGAGGAGGAGGACCACTGCTCCCGAGAACCCACCTGCAGGCA 219  
Db 2263 CGGGGCGGAAGCGGAGGCTGAGCGCTCTCGGGCCCCCTCGGGGTGGGA 2312

Search completed: February 22, 2005, 00:06:48  
Job time : 166 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 23:10:41 ; Search time 500 Seconds  
(without alignments)  
7553.614 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 639

Sequence: 1 atcgctgcctccgcacg.....ataaccaggtgaccatcaag 639

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5384158 seqs, 2955248155 residues

Total number of hits satisfying chosen parameters: 10768316

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	261.8	41.0	497	9 US-09-864-761-156	Sequence 156, App
C 2	239.8	37.5	243	9 US-09-864-761-16984	Sequence 16984, A
C 3	91.8	14.4	1128	18 US-10-363-345A-17159	Sequence 17159, A
C 4	91.8	14.4	1128	18 US-10-363-345A-17160	Sequence 17160, A
C 5	85.6	13.4	1128	18 US-10-363-345A-17157	Sequence 17157, A
C 6	85.6	13.4	1128	18 US-10-363-345A-17158	Sequence 17158, A
7	46.2	7.2	1770	17 US-10-425-114-25086	Sequence 25086, A
8	46.2	7.2	1770	18 US-10-425-115-165209	Sequence 165209, A
9	42.6	6.7	3018	9 US-09-802-127-6	Sequence 6, Appli
10	42.6	6.7	3391	9 US-09-802-127-4	Sequence 4, Appli
11	42.6	6.7	5968	9 US-09-764-868-1501	Sequence 1501, Ap

12	42.6	6.7	5968	14	US-10-103-313-619	Sequence 619, App
13	42.4	6.6	3368	18	US-10-437-963-13512	Sequence 13512, A
C 14	42.2	6.6	815	9	US-09-823-038A-9	Sequence 9, Appli
15	42	6.6	738	18	US-10-425-115-163054	Sequence 163054, A
16	41.6	6.5	149430	18	US-10-626-832-1	Sequence 1, Appli
C 17	41.6	6.5	149430	18	US-10-626-832-1	Sequence 1, Appli
C 18	41.6	6.5	150223	18	US-10-624-149A-1	Sequence 1, Appli
C 19	41.6	6.5	150223	18	US-10-624-149A-1	Sequence 1, Appli
20	41.4	6.5	537	18	US-10-437-963-66677	Sequence 66677, A
21	40.8	6.4	442	18	US-10-425-115-88722	Sequence 88722, A
C 22	40.8	6.4	1259	18	US-10-437-963-100621	Sequence 100621, A
C 23	40.8	6.4	1409	18	US-10-437-963-52890	Sequence 52890, A
C 24	40.6	6.4	101954	18	US-10-322-281-782	Sequence 782, App
25	40.6	6.4	200000	18	US-10-672-764A-31	Sequence 31, Appli
26	40.4	6.3	1732	18	US-10-437-963-25745	Sequence 25745, A
27	40.4	6.3	2000	17	US-10-260-238-23359	Sequence 2359, Ap
28	40.4	6.3	4426	17	US-10-366-547-46	Sequence 46, Appli
29	40.2	6.3	42339	13	US-10-087-192-991	Sequence 991, App
30	40	6.3	605	11	US-09-764-875-117	Sequence 117, App
C 31	39.8	6.2	345	18	US-10-425-115-94260	Sequence 94260, A
32	39.8	6.2	831	18	US-10-437-963-66678	Sequence 66678, A
33	39.8	6.2	981	17	US-10-425-114-22196	Sequence 22196, A
34	39.8	6.2	981	18	US-10-425-115-94549	Sequence 94549, A
35	39.8	6.2	15872	9	US-09-861-289-1	Sequence 1, Appli
36	39.8	6.2	15872	9	US-09-860-846-1	Sequence 1, Appli
37	39.8	6.2	15872	10	US-09-988-384B-1	Sequence 1, Appli
38	39.8	6.2	15872	10	US-09-836-821-1	Sequence 1, Appli
39	39.8	6.2	15872	16	US-10-271-889-44	Sequence 44, Appli
40	39.6	6.2	516	18	US-10-437-963-48622	Sequence 48622, A
41	39.6	6.2	1261	18	US-10-437-963-70724	Sequence 70724, A
42	39.4	6.2	653	13	US-10-027-632-132758	Sequence 132758, A
43	39.4	6.2	653	13	US-10-027-632-132759	Sequence 132759, A
44	39.4	6.2	653	13	US-10-027-632-132760	Sequence 132760, A
45	39.4	6.2	653	17	US-10-027-632-132758	Sequence 132758, A

#### ALIGNMENTS

#### RESULT 1

US-09-864-761-156/c Application US/09864761  
; Sequence 156, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30





RESULT 4  
US-10-363-345A-17160/c  
; Sequence 17160, Application US/10363345A  
; Publication No. US20040234960A1

Query Match	13.4%	Score 85.6;	DB 18;	Length 1128;
Best Local Similarity	75.7%;	Pred. No. 5.7e-14;		
Matches 106; Conservative	0;	Mismatches 34;	Indels 0;	Gaps 0;

  

Qy	1	A T G C G T G C G T C C G C A C C G A G C G G G C T C T C T C T G T G T G T G T G C G C G C G C G T G	60
Db	521	A T A C G T A C G T C T C G C A C C G A C C G A A C T C C T C T A C G T A C T A C T A A C A C G C T A	462
Qy	61	C T G A G C G C G C G T A G G C T C C C G T G A A G A G C C G C G G T C C C G C A C A C G C G C T G G G	120
Db	461	C T A A A A C A C G C T A A A A C T C C C G T A A A A A A C C G C G A C T C C C G A A C C A C G A C C T A A A	402
Qy	121	A G C C T A C A G G C T C G C A G A	140







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 22:00:30 ; Search time 3098 Seconds  
(without alignments)  
7851.216 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696

Perfect score: 639  
Sequence: 1 atcgctgcgtccgcgacgcg.....ataaccaggtgaccatcaag 639

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gses1: \*  
9: gb\_gses2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	637.4	99.7	790	5	BX336289
2	637.4	99.7	826	5	BX387234
3	637.4	99.7	866	5	BX365974
4	637.4	99.7	867	5	BX403263
5	637.4	99.7	868	5	BX365998
6	637.4	99.7	872	5	BX344975
7	637.4	99.7	900	5	BX398022
8	637.4	99.7	917	5	BX334584
9	637.4	99.7	936	5	BX357282
10	637.4	99.7	936	5	BX381620
11	637.4	99.7	947	5	BX359737
12	637.4	99.7	960	5	BX381477
13	637.4	99.7	968	5	BX356306
14	637.4	99.7	971	5	BX336623
15	637.4	99.7	973	5	BX379676
16	637.4	99.7	975	5	BX382082
17	637.4	99.7	978	5	BX336717
18	637.4	99.7	979	5	BX335053
19	637.4	99.7	979	5	BX360184
20	637.4	99.7	980	5	BX360878
21	637.4	99.7	983	1	AL554329
22	637.4	99.7	983	5	BX358077
23	637.4	99.7	983	5	BX361532
24	637.4	99.7	985	5	BX458619

25	637.4	99.7	994	5	BX339696
26	637.4	99.7	3009	3	CR749863
27	637	99.7	782	5	BX360064
28	637	99.7	967	5	BX357315
29	636.6	99.6	942	5	BX344130
30	636.4	99.6	881	5	BX345009
31	636	99.5	984	5	BX337459
32	635.8	99.5	882	5	BX387069
33	635.8	99.5	924	5	BX378610
34	635.8	99.5	955	5	BX343795
35	635.8	99.5	985	3	CNSLT11BO
36	635	99.4	708	5	BX324767
37	635	99.4	960	1	AL549276
38	634.2	99.2	870	5	BX387067
39	631.6	98.8	868	5	BX366007
40	630.4	98.7	938	5	BX397246
41	629.8	98.6	913	5	BX366004
42	629.4	98.5	967	5	BX35752
43	629.4	98.5	1113	5	BX439514
44	626.4	98.0	982	5	BX359034
45	624.8	97.8	982	5	BX358526

ALIGNMENTS

RESULT 1  
LOCUS BX336289 790 bp mRNA linear EST 07-APR-2004  
DEFINITION BX336289 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA clone CSODI027YA01 5-PRIME, mRNA sequence.  
ACCESSION BX336289  
VERSION BX336289.2 GI:46274766  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 790)  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 2, 2003 this sequence version replaced gi:30341500.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r  
There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?s=CSODI027AA01QP1&c=7348.r>.

FEATURES

Location/Qualifiers  
1..790  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI027YA01"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 790;  
Best Local Similarity 99.8%; Pred. No. 5.4e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?s=CS5AA0142D04RM1&c=7348.r>.

FEATURES  
source  
1. 826  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DI034YP24"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 826;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGTGGCTCCCGGACCGAGCCGGGCTCTCTCTGCGTGTGCTGCTGGCGGCGCTG 60  
DB 15 ATGCGTGGCTCCCGGACCGAGCCGGGCTCTCTCTGCGTGTGCTGCTGGCGGCGCTG 74  
QY 61 CTGAGGCGGCGCTAGGCTCCCGTGAAGAAGCGCGGCTCCCGGACCAACGCGCTGGG 120  
DB 75 CTGAGGCGGCGCTAGGCTCCCGTGAAGAAGCGCGGCTCCCGGACCAACGCGCTGGG 134  
QY 121 AGCTCTACGAGGCTCGCAGAGGCTCTAGCCTCCAGATCTAGGCTCTGAAGAAGAG 180  
DB 135 AGCTCTACGAGGCTCGCAGAGGCTCTAGCCTCCAGATCTAGGCTCTGAAGAAGAG 194  
QY 181 GAGAGGACACATGCTCTCCAGAACCCAGCTGAGGAGGACCAACCAACATGATGC 240  
DB 195 GAGAGGACACATGCTCTCCAGAACCCAGCTGAGGAGGACCAACCAACATGATGC 254  
QY 241 TGGAGTGTCACTGAGCAGAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 300  
DB 255 TGGAGTGTCACTGAGCAGAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 314  
QY 301 GAGTTACTCGTTCGCGCTGGAGTGCAGAGCTGCGGAGTGGCGAGCACAACCTTG 360  
DB 315 GAGTTACTCGTTCGCGCTGGAGTGCAGAGCTGCGGAGTGGCGAGCACAACCTTG 374  
QY 361 AGTACCCCTAACCTGATACCCAGGCTTACGCTCCCGAGATCTAGGCTCTGAGGAA 420  
DB 375 AGTACCCCTAACCTGATACCCAGGCTTACGCTCCCGAGATCTAGGCTCTGAGGAA 434  
QY 421 GAGGAGGAGGACGACTCTCTCCAGAACCCAGCTGAGGAGGAGTACACCAATGGA 480  
DB 435 GAGGAGGAGGACGACTCTCTCCAGAACCCAGCTGAGGAGGAGTACACCAATGGA 494  
QY 481 TGTGGAGTGTCACTGAGCAGAGCCTGACCCAGGAGTGCAGAGGATGCCAGCCTCCAGGACC 540  
DB 495 TGTGGAGTGTCACTGAGCAGAGCCTGACCCAGGAGTGCAGAGGATGCCAGCCTCCAGGACC 554  
QY 541 CAGGAGGTTACTCCCTTCTGCTGAGCTGAGAGCTGCCAGATTTGGTCCAGCAACC 600  
DB 555 CAGGAGGTTACTCCCTTCTGCTGAGCTGAGAGCTGCCAGATTTGGTCCAGCAACC 614  
QY 601 TTGAGTACCCCTAACCTGATACCCAGGAGTGCAGGATCAAG 639  
DB 615 TTGAGTACCCCTAACCTGATACCCAGGAGTGCAGGATCAAG 653

RESULT 2  
BX387234  
LOCUS  
DEFINITION  
BX387234 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0DI034YP24 5-PRIME, mRNA sequence.  
BX387234  
BX387234.2 GI:46571918  
EST.  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1. (bases 1 to 826)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 8, 2003 this sequence version replaced gi:30447511.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r

RESULT 3  
BX365974  
LOCUS  
DEFINITION  
BX365974 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0DI031YB04 5-PRIME, mRNA sequence.  
BX365974  
BX365974.2 GI:46288863  
ACCESSION  
VERSION  
BX365974  
BX365974.2  
GI:46288863

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?s=CS5AA0142D04RM1&c=7348.r>.

FEATURES  
source  
1. 826  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DI034YP24"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 826;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGTGGCTCCCGGACCGAGCCGGGCTCTCTCTGCGTGTGCTGCTGGCGGCGCTG 60  
DB 5 ATGCGTGGCTCCCGGACCGAGCCGGGCTCTCTCTGCGTGTGCTGCTGGCGGCGCTG 64  
QY 61 CTGAGGCGGCGCTAGGCTCCCGTGAAGAAGCGCGGCTCCCGGACCAACGCGCTGGG 120  
DB 65 CTGAGGCGGCGCTAGGCTCCCGTGAAGAAGCGCGGCTCCCGGACCAACGCGCTGGG 124  
QY 121 AGCTCTACGAGGCTCGCAGAGGCTCTAGCCTCCAGATCTAGGCTCTGAAGAAGAG 180  
DB 125 AGCTCTACGAGGCTCGCAGAGGCTCTAGCCTCCAGATCTAGGCTCTGAAGAAGAG 184  
QY 181 GAGGAGGACACATGCTCTCCAGAACCCAGCTGAGGAGGAGTACACCAATGGA 240  
DB 185 GAGGAGGACACATGCTCTCCAGAACCCAGCTGAGGAGGAGTACACCAATGGA 244  
QY 241 TGGAGTGTCACTGAGCAGAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 300  
DB 245 TGGAGTGTCACTGAGCAGAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 304  
QY 301 GAGTTACTCGTTCGCGCTGGAGTGCAGAGCTGCGGAGTGGCGAGCACAACCTTG 360  
DB 305 GAGTTACTCGTTCGCGCTGGAGTGCAGAGCTGCGGAGTGGCGAGCACAACCTTG 364  
QY 361 AGTACCCCTAACCTGATACCCAGGCTTACGCTCCCGAGATCTAGGCTCTGAGGAA 420  
DB 365 AGTACCCCTAACCTGATACCCAGGCTTACGCTCCCGAGATCTAGGCTCTGAGGAA 424  
QY 421 GAGGAGGAGGACGACTCTCTCCAGAACCCAGCTGAGGAGGAGTACACCAATGGA 480  
DB 425 GAGGAGGAGGACGACTCTCTCCAGAACCCAGCTGAGGAGGAGTACACCAATGGA 484  
QY 481 TGTGGAGTGTCACTGAGCAGAGCCTGACCCAGGAGTGCAGAGGATGCCAGCCTCCAGGACC 540  
DB 485 TGTGGAGTGTCACTGAGCAGAGCCTGACCCAGGAGTGCAGAGGATGCCAGCCTCCAGGACC 544  
QY 541 CAGGAGGTTACTCCCTTCTGCTGAGCTGAGAGCTGCCAGATTTGGTCCAGCAACC 600  
DB 545 CAGGAGGTTACTCCCTTCTGCTGAGCTGAGAGCTGCCAGATTTGGTCCAGCAACC 604  
QY 601 TTGAGTACCCCTAACCTGATACCCAGGAGTGCAGGATCAAG 639  
DB 605 TTGAGTACCCCTAACCTGATACCCAGGAGTGCAGGATCAAG 643

RESULT 3  
BX365974  
LOCUS  
DEFINITION  
BX365974 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0DI031YB04 5-PRIME, mRNA sequence.  
BX365974  
BX365974.2  
GI:46288863



```
Qy 181 GAGGAGGACCACTGCTCTCCAGAACCCACCTGCAGGCGAGGACCAACCAATGGATGC 240
Db 208 GAGGAGGACCACTGCTCTCCAGAACCCACCTGCAGGCGAGGACCAACCAATGGATGC 267
Qy 241 TGGACTGTCACTGAGCAGCAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 300
Db 268 TGGACTGTCACTGAGCAGCAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 327
Qy 301 GAGGTTACTTCGTTGCGGCTGAGCTGCAGAGCTGCCGGGATTGGCCAGCACCAACCTTG 360
Db 328 GAGGTTACTTCGTTGCGGCTGAGCTGCAGAGCTGCCGGGATTGGCCAGCACCAACCTTG 387
Qy 361 AGTACCCCTAACCCCTGATACCCAGGCTTCAGCTCCCCAGATCCTAGGCGCTTCTGAGGGAA 420
Db 388 AGTACCCCTAACCCCTGATACCCAGGCTTCAGCTCCCCAGATCCTAGGCGCTTCTGAGGGAA 447
Qy 421 GAGGAGGAGGACGACTGCTCTCCAGAACCCACCTGCAGGCGAGGACCAACCAATGGGA 480
Db 448 GAGGAGGAGGACGACTGCTCTCCAGAACCCACCTGCAGGCGAGGACCAACCAATGGGA 507
Qy 481 TGTGGACTGTCACTGAGCAGCAGCCTGACCCAGGGAATGCCAGCCTCCAGGACCC 540
Db 508 TGTGGACTGTCACTGAGCAGCAGCCTGACCCAGGGAATGCCAGCCTCCAGGACCC 567
Qy 541 CAGGAGGTTACTTCCTTCTGCTGAGCTGCAGAGCTGCCAGAAATTTGGTCCACGCAACC 600
Db 568 CAGGAGGTTACTTCCTTCTGCTGAGCTGCAGAGCTGCCAGAAATTTGGTCCACGCAACC 627
Qy 601 TTGAGTACCCCTAACCCCTGATAACCAAGTGACCATCAAG 639
Db 628 TTGAGTACCCCTAACCCCTGATAACCAAGTGACCATCAAG 666
```

```
RESULT 5
BX365998
LOCUS
DEFINITION
BX365998 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1074Y001 5-PRIME, mRNA sequence.
ACCESSION
BX365998
VERSION
BX365998.2 GI:46306893
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 668)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 5, 2003 this sequence version replaced gi:30376966.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7348.r
```

```
There is a virtual cDNA representing this cluster. For more
information about this cluster and the virtual cDNA, see
http://www.genoscope.cns.fr/cdna?s=CS2BAX142D10_AX242B5_1&c=7348.r.
Location/Qualifiers
1..868
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1074Y001"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
```

```
ORIGIN
Query Match 99.7%; Score 637.4; DB 5; Length 868;
Best Local Similarity 99.8%; Pred. No. 5.5e-135;
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCGTGGCTCCCGGACCGAGCCGGGCTCTCTCTGCGTGTGCTGCTGCGCGCGCTG 60
Db 7 ATGCGTGGCTCCCGGACCGAGCCGGGCTCTCTCTGCGTGTGCTGCTGCGCGCGCTG 66
Qy 61 CTGAGGCGGCGCTAGGCGCTCCCGTGAAGAACCGCGGCTCCGCGAGCACGCGCTGG 120
Db 67 CTGAGGCGGCGCTAGGCGCTCCCGTGAAGAACCGCGGCTCCGCGAGCACGCGCTGG 126
Qy 121 AGCCTCAGAGGCTCGCAGAGGCTCTCAGCCTCCCGAGATCCTAGGCTCTCAAGGAAG 180
Db 127 AGCCTCAGAGGCTCGCAGAGGCTCTCAGCCTCCCGAGATCCTAGGCTCTCAAGGAAG 186
Qy 181 GAGGAGGACCACTGCTCCCGAGAACCCACCTGCAGGCGAGGACCAACCAATGGATGC 240
Db 187 GAGGAGGACCACTGCTCCCGAGAACCCACCTGCAGGCGAGGACCAACCAATGGATGC 246
Qy 241 TGGACTGTCACTGAGCAGCAGCCTGACCCAGGCAACACACCCCTCCAGGACCCCA 300
Db 247 TGGACTGTCACTGAGCAGCAGCCTGACCCAGGCAACACACCCCTCCAGGACCCCA 306
Qy 301 GAGGTTACTTCGTTGCGGCTGAGCTGCAGAGCTGCCGGGATTGGCCAGCACCAACCTTG 360
Db 307 GAGGTTACTTCGTTGCGGCTGAGCTGCAGAGCTGCCGGGATTGGCCAGCACCAACCTTG 366
Qy 361 AGTACCCCTAACCCCTGATAACCCAGGCTTCAGCCTCCCGAGATCCTAGGCTCTGAGGGAA 420
Db 367 AGTACCCCTAACCCCTGATAACCCAGGCTTCAGCCTCCCGAGATCCTAGGCTCTGAGGGAA 426
Qy 421 GAGGAGGAGGACGACTGCTCTCCAGAACCCACCTGCAGGCGAGGCTACCAACATGGA 480
Db 427 GAGGAGGAGGACGACTGCTCTCCAGAACCCACCTGCAGGCGAGGCTACCAACATGGA 486
Qy 481 TGTGGACTGTCACTGAGCAGCAGCCTGACCCAGGGAATGCCAGCCTCCAGGACCC 540
Db 487 TGTGGACTGTCACTGAGCAGCAGCCTGACCCAGGGAATGCCAGCCTCCAGGACCC 546
Qy 541 CAGGAGGTTACTTCCTTCTGCTGAGCTGCAGAGCTGCCAGAAATTTGGTCCACGCAACC 600
Db 547 CAGGAGGTTACTTCCTTCTGCTGAGCTGCAGAGCTGCCAGAAATTTGGTCCACGCAACC 606
Qy 601 TTGAGTACCCCTAACCCCTGATAACCAAGTGACCATCAAG 639
Db 607 TTGAGTACCCCTAACCCCTGATAACCAAGTGACCATCAAG 645
```

RESULT 6  
BX344975

LOCUS  
DEFINITION  
BX344975 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0D1005YF06 5-PRIME, mRNA sequence.

ACCESSION  
BX344975  
VERSION  
BX344975.2 GI:46282437

KEYWORDS  
EST.

SOURCE  
Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 872)

Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 2, 2003 this sequence version replaced gi:30346327.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE



Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see [http://www.genoscope.cns.fr/cdna?s=CS2BAX12ZA01\\_AX19ZH11\\_l&c=7348.r](http://www.genoscope.cns.fr/cdna?s=CS2BAX12ZA01_AX19ZH11_l&c=7348.r)

```
FEATURES
  source
    Location/Qualifiers
      1..872
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0D1005YP06"
        /tissue_type="PLACENTA COT 25-NORMALIZED"
        /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
        /notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
      ORIGIN
        Query Match      99.7%; Score 637.4; DB 5; Length 872;
        Best Local Similarity 99.8%; Pred. No. 5.5e-135;
        Mismatches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGTGGCTCCGCGACCGCGGCTCTCTCTGCGTGGCTGGCGGCGCTG 60
Db 13 ATGCGTGGCTCCGCGACCGCGGCTCTCTCTGCGTGGCTGGCGGCGCTG 72

Qy 61 CTGAGGCGGCGCTAGGCTCTCCGTAAGAGCGCGGCTCCGCGACCAACGCGCTGG 120
Db 73 CTGAGGCGGCGCTAGGCTCTCCGTAAGAGCGCGGCTCCGCGACCAACGCGCTGG 132

Qy 121 AGCTCTACGAGGCTCGCAGAGGCTCTAGCTCTCCAGATCTTAGGCTCTGAAGGAAG 180
Db 133 AGCTCTACGAGGCTCGCAGAGGCTCTAGCTCTCCAGATCTTAGGCTCTGAAGGAAG 192

Qy 181 GAGGAGGACCACTCTCCAGAGACCACTCTGAGGAGGAGGACCAACATGGATGC 240
Db 193 GAGGAGGACCACTCTCCAGAGACCACTCTGAGGAGGAGGACCAACATGGATGC 252

Qy 241 TGGAGTGTCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db 253 TGGAGTGTCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 312

Qy 301 GAGGTTACTCCGTTGCGGCTGGAGCTGAGAGGCTGCGGAGGATGGCCACCAACCTTG 360
Db 313 GAGGTTACTCCGTTGCGGCTGGAGCTGAGAGGCTGCGGAGGATGGCCACCAACCTTG 372

Qy 361 AGTACCCCTAACCTGATACCGAGGCTTACGCTTCCGAGATCTTAGGCTCTGAGGAG 420
Db 373 AGTACCCCTAACCTGATACCGAGGCTTACGCTTCCGAGATCTTAGGCTCTGAGGAG 432

Qy 421 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
Db 433 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 492

Qy 481 TGTGAGTGTCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Db 493 TGTGAGTGTCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552

Qy 541 CAGAGGTTACTCCCTTGTCTGGAGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGG 600
Db 553 CAGAGGTTACTCCCTTGTCTGGAGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGG 612

Qy 601 TTGAGTACCCCTAACCTGATACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 639
Db 613 TTGAGTACCCCTAACCTGATACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 651
```

RESULT 7  
BX398022  
LOCUS  
DEFINITION  
BX398022 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0D10051YH09 5-PRIME, mRNA sequence.  
BX398022  
ACCESSION  
BX398022.2 GI:46876522  
VERSION  
BX398022  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1. (bases 1 to 900)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
On May 13, 2003 this sequence version replaced gi:30621466.  
COMMENT  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see [http://www.genoscope.cns.fr/cdna?s=CS0D10051YH09\\_l&c=7348.r](http://www.genoscope.cns.fr/cdna?s=CS0D10051YH09_l&c=7348.r).

FEATURES  
 source

1..900  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D10051YH09"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 900;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Mismatches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 1 ATGCGTGGCTCCGCGACCGCGGCTCTCTCTGCGTGGCTGGCGGCGCTG 60
Db 13 ATGCGTGGCTCCGCGACCGCGGCTCTCTCTGCGTGGCTGGCGGCGCTG 72

Qy 61 CTGAGGCGGCGCTAGGCTCTCCGTAAGAGCGCGGCTCCGCGACCAACGCGCTGG 120
Db 73 CTGAGGCGGCGCTAGGCTCTCCGTAAGAGCGCGGCTCCGCGACCAACGCGCTGG 132

Qy 121 AGCTCTACGAGGCTCGCAGAGGCTCTAGCTCTCCAGATCTTAGGCTCTGAAGGAAG 180
Db 133 AGCTCTACGAGGCTCGCAGAGGCTCTAGCTCTCCAGATCTTAGGCTCTGAAGGAAG 192

Qy 181 GAGGAGGACCACTCTCCAGAGACCACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
Db 193 GAGGAGGACCACTCTCCAGAGACCACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 252

Qy 241 TGGAGTGTCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
Db 253 TGGAGTGTCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 312

Qy 301 GAGGTTACTCCGTTGCGGCTGGAGCTGAGAGGCTGCGGAGGATGGCCACCAACCTTG 360
Db 313 GAGGTTACTCCGTTGCGGCTGGAGCTGAGAGGCTGCGGAGGATGGCCACCAACCTTG 372
```



```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI023YN23"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      99.7%; Score 637.4; DB 5; Length 936;
Best Local Similarity 99.8%; Pred. No. 5.5e-135;
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTCTGCTGCGCGCGCTG 60
Db 9 ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTCTGCTGCGCGCGCTG 68
Qy 61 CTGAGGCGGGCTAGGGCTCCCGTGAAGAGCGGGCTCCGCGACCAACGCCCTGGG 120
Db 69 CTGAGGCGGGCTAGGGCTCCCGTGAAGAGCGGGCTCCGCGACCAACGCCCTGGG 128
Qy 121 AGCTCAGAGGCTCGCAGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGAGAG 180
Db 129 AGCTCAGAGGCTCGCAGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGAGAG 188
Qy 181 GAGGAGGACCACTGCTCCCGAGAACCCCTGCGAGGAGCCACACCAATGGATGC 240
Db 189 GAGGAGGACCACTGCTCCCGAGAACCCCTGCGAGGAGCCACACCAATGGATGC 248
Qy 241 TGGACTGTCTAGGCGAGCGCCATGACCCGAGCAACACACCCCTCCCGAGACCCA 300
Db 249 TGGACTGTCTAGGCGAGCGCCATGACCCGAGCAACACACCCCTCCCGAGACCCA 308
Qy 301 GAGGTTACTCGTGGCTGGAGCTGAGAGCTGCGGGATGCGGGATGCGGCAACAC 360
Db 309 GAGGTTACTCGTGGCTGGAGCTGAGAGCTGCGGGATGCGGGATGCGGCAACAC 368
Qy 361 AGTACCCCTAACCTTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAG 420
Db 369 AGTACCCCTAACCTTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAG 428
Qy 421 GAGGAGGAGGACGACTGCTCCCGAGAACCCCTGCGAGGAGCTACACCAATGGA 480
Db 429 GAGGAGGAGGACGACTGCTCCCGAGAACCCCTGCGAGGAGCTACACCAATGGA 488
Qy 481 TGTGGACTGTCTAGGCGAGCGCTGACCCGAGGATGCGGCTCCCGAGGACC 540
Db 489 TGTGGACTGTCTAGGCGAGCGCTGACCCGAGGATGCGGCTCCCGAGGACC 548
Qy 541 CAGGAGGTTACTCCCTTGCTGCTGAGCTGAGAGCTGCCAGAAATTGGTCCAGCA 600
Db 549 CAGGAGGTTACTCCCTTGCTGCTGAGCTGAGAGCTGCCAGAAATTGGTCCAGCA 608
Qy 601 TTGAGTACCCCTAACCTTGATACCCAGGCTGACCATCAAG 639
Db 609 TTGAGTACCCCTAACCTTGATACCCAGGTTGACCATCAAG 647

RESULT 10
LOCUS BX381620
DEFINITION BX381620 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
          clone CS0DI068YE01 5-PRIME, mRNA sequence.
ACCESSION BX381620
VERSION BX381620.2 GI:46570758
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 936)
```

AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 8, 2003 this sequence version replaced gi:30443700.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and EcoR V  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r

There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
http://www.genoscope.cns.fr/cdna?s=CS0DI068AC01Q1&c=7348.r.

FEATURES  
Location/Qualifiers  
1. 936  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DI068YE01"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/notes="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector.  
Library was normalized."

ORIGIN  
Query Match 99.7%; Score 637.4; DB 5; Length 936;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTCTGCTGCGCGCGCTG 60  
Db 13 ATGCGTGGCTCCGCGACCGAGCGGGCTCTCTCTGCGTGTCTGCTGCGCGCGCTG 72  
Qy 61 CTGAGGCGGGCTAGGGCTCCCGTGAAGAGCGGGCTCCGCGACCAACGCCCTGGG 120  
Db 73 CTGAGGCGGGCTAGGGCTCCCGTGAAGAGCGGGCTCCGCGACCAACGCCCTGGG 132  
Qy 121 AGCTCAGAGGCTCGCAGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGAGAG 180  
Db 133 AGCTCAGAGGCTCGCAGAGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGAGAG 192  
Qy 181 GAGGAGGACCACTGCTCCCGAGAACCCCTGCGAGGAGCCACACCAATGGATGC 240  
Db 193 GAGGAGGACCACTGCTCCCGAGAACCCCTGCGAGGAGCCACACCAATGGATGC 252  
Qy 241 TGGACTGTCTAGGCGAGCGCCATGACCCGAGCAACACACCCCTCCCGAGACCCA 300  
Db 253 TGGACTGTCTAGGCGAGCGCCATGACCCGAGCAACACACCCCTCCCGAGACCCA 312  
Qy 301 GAGGTTACTCGTGGCTGGAGCTGAGAGCTGCGGGATGCGGGATGCGGCAACAC 360  
Db 313 GAGGTTACTCGTGGCTGGAGCTGAGAGCTGCGGGATGCGGGATGCGGCAACAC 372  
Qy 361 AGTACCCCTAACCTTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAG 420  
Db 373 AGTACCCCTAACCTTGATACCCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAG 432  
Qy 421 GAGGAGGAGGACGACTGCTCCCGAGAACCCCTGCGAGGAGCTACACCAATGGA 480  
Db 433 GAGGAGGAGGACGACTGCTCCCGAGAACCCCTGCGAGGAGCTACACCAATGGA 492  
Qy 481 TGTGGACTGTCTAGGCGAGCGCTGACCCGAGGATGCGGCTCCCGAGGACC 540  
Db 493 TGTGGACTGTCTAGGCGAGCGCTGACCCGAGGATGCGGCTCCCGAGGACC 552  
Qy 541 CAGGAGGTTACTCCCTTGCTGCTGAGCTGAGAGCTGCCAGAAATTGGTCCAGCA 600

```
Db 553 CAGGAGGTACTCCCTGCTGCTGAGAGCTGCAGAAAGTGCAGATTTGGTCCACGCAACC 612
Qy 601 TTGAGTACCCCTAACCCCTGATACACAGGTGACCATCAAG 639
Db 613 TTGAGTACCCCTAACCCCTGATACACAGGTGACCATCAAG 651

RESULT 11
LOCUS BX359737
DEFINITION BX359737 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI060YF04 5-PRIME, mRNA sequence.
ACCESSION BX359737
VERSION BX359737.2 GI:46305712
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 947)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 5, 2003 this sequence version replaced gi:30376366.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7348.r
There is a virtual cDNA representing this cluster. For more
information about this cluster and the virtual cDNA, see
http://www.genoscope.cns.fr/cdna?s=CSODI060DC02QPI&c=7348.r.
Location/Qualifiers
1. 947
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI060YF04"
/tissue type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notice="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 99.7%; Score 637.4; DB 5; Length 947;
Best Local Similarity 99.8%; Pred. No. 5.5e-135;
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGTGCGCTCCGCGACGAGCGGGCTCTCTCTCGTGCTGCTGCGCGCGCTG 60
Db 14 ATGCGTGCGCTCCGCGACGAGCGGGCTCTCTCTCGTGCTGCTGCGCGCGCTG 73

Qy 61 CTGAGGCGGCGCTAGGGCTCCCGTGAAGAAGCGCGGCTCCGCGGACCAACGCGCTGG 120
Db 74 CTGAGGCGGCGCTAGGGCTCCCGTGAAGAAGCGCGGCTCCGCGGACCAACGCGCTGG 133

Qy 121 AGCTCTCAGAGGCTCGCAGAGGTCTCAGCTCCCGAGTCTTAGGCTCTGAAGAGAG 180
Db 134 AGCTCTCAGAGGCTCGCAGAGGTCTCAGCTCCCGAGTCTTAGGCTCTGAAGAGAG 193

Qy 181 GAGGAGGACCACTGCTCTCCCGAGAACCCACCTGCGAGGACGACACCAACATGGATGC 240
Db 194 GAGGAGGACCACTGCTCTCCCGAGAACCCACCTGCGAGGACGACACCAACATGGATGC 253

Qy 241 TGGAGTGTACTGAGCGAGCGGATGATGACCCCGAGCAACACCCCTCCAGGACCCCA 300
```

```
Db 254 TGGACTGTCTAGCGAGCAGCAGCCATGACCCCGAGCAACACCCCTCCAGGACCCCA 313
Qy 301 GAGTTTACTTCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATTGGCCAGACAACTTGT 360
Db 314 GAGTTTACTTCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATTGGCCAAACACAACTTGT 373
Qy 361 AGTACCCCTAACCCCTGATACCCAGGCTTACGCTCCCGAGATCTCTAGGCGCTCTGAGGAA 420
Db 374 AGTACCCCTAACCCCTGATACCCAGGCTTACGCTCCCGAGATCTCTAGGCGCTCTGAGGAA 433
Qy 421 GAGGAGGAGGACGACTGTCTCCCGAGAACCCACCTGCGAGGACGAGCTACACCAATGGA 480
Db 434 GAGGAGGAGGACGACTGTCTCCCGAGAACCCACCTGCGAGGACGAGCTACACCAATGGA 493
Qy 481 TGTGGAGTGTCTAGCTGAGCGAGCGCTGACCCCGAGGATGCCAGCGCTCCAGGACC 540
Db 494 TGTGGAGTGTCTAGCTGAGCGAGCGCTGACCCCGAGGATGCCAGCGCTCCAGGACC 553
Qy 541 CAGGAGGTCTACTCCCTTGTCTGAGCTGCAGAGCTGCAGAGTGGTCCAGCGCAACC 600
Db 554 CAGGAGGTCTACTCCCTTGTCTGAGCTGCAGAGCTGCAGAGTGGTCCAGCGCAACC 613
Qy 601 TTGAGTACCCCTAACCCCTGATACCCAGGCTGACCATCAAG 639
Db 614 TTGAGTACCCCTAACCCCTGATACCCAGGCTGACCATCAAG 652

RESULT 12
LOCUS BX381477
DEFINITION BX381477 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI066YK17 5-PRIME, mRNA sequence.
ACCESSION BX381477
VERSION BX381477.2 GI:46622135
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 960)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 8, 2003 this sequence version replaced gi:30452951.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7348.r
There is a virtual cDNA representing this cluster. For more
information about this cluster and the virtual cDNA, see
http://www.genoscope.cns.fr/cdna?s=CSODI066AF09QPI&c=7348.r.
Location/Qualifiers
1. 960
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI066YK17"
/tissue type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notice="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 99.7%; Score 637.4; DB 5; Length 960;
Best Local Similarity 99.8%; Pred. No. 5.5e-135;
```

Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGTGGCTCCGCGACGAGCGGGCTCTCTCTCGTCTGCTCTGCGCGCGCTG 60  
DB 14 ATGCGTGGCTCCGCGACGAGCGGGCTCTCTCTCGTCTGCTCTGCGCGCGCTG 73  
QY 61 CTGAGGCGGGCTAGGGCTCCCGTGAAGAGCGCGCTCCGCGACCAACGCGCTGGG 120  
DB 74 CTGAGGCGGGCTAGGGCTCCCGTGAAGAGCGCGCTCCGCGACCAACGCGCTGGG 133  
QY 121 AGCTCTCAGAGGCTCGAGAGTCTCAGCTCCCGAGATCTAGGCTCTGAAGAGAG 180  
DB 134 AGCTCTCAGAGGCTCGAGAGTCTCAGCTCCCGAGATCTAGGCTCTGAAGAGAG 193  
QY 181 GAGGAGGACCACTGCTCTCCAGAACCCCTGAGGAGGAGCCACCAACATGATGC 240  
DB 194 GAGGAGGACCACTGCTCTCCAGAACCCCTGAGGAGGAGCCACCAACATGATGC 253  
QY 241 TGGAGTGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
DB 254 TGGAGTGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 313  
QY 301 GAGGTTACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360  
DB 314 GAGGTTACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 373  
QY 361 AGTACCCCTTAACCTGATACCCAGGCTTCAGGCTCCCGAGATCTAGGCTCTGAGGAA 420  
DB 374 AGTACCCCTTAACCTGATACCCAGGCTTCAGGCTCCCGAGATCTAGGCTCTGAGGAA 433  
QY 421 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480  
DB 434 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 493  
QY 481 TGTGGAGTGTACTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
DB 494 TGTGGAGTGTACTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 553  
QY 541 CAGGAGGTTACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600  
DB 554 CAGGAGGTTACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 613  
QY 601 TTGAGTACCCCTTAACCTGATACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 639  
DB 614 TTGAGTACCCCTTAACCTGATACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 652

## RESULT 13

BX356306  
LOCUS  
DEFINITION BX356306 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CSODI010YD02 5-PRIME, mRNA sequence.  
BX356306  
ACCESSION  
VERSION BX356306.2 GI:46303698  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

## REFERENCE

1. (Bases 1 to 968)  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30374043.

## COMMENT

Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with NotI and cloned  
into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
<http://www.genoscope.cns.fr/cdna?s=CSODI010YD01QPI&c=7348.r>.

## FEATURES

source

1. .968  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI010YD02"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with NotI and cloned into the NotI and EcoRV  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 99.7%; Score 637.4; DB 5; Length 968;  
Best Local Similarity 99.8%; Pred. No. 5.5e-135;  
Matches 638; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGTGGCTCCGCGACGAGCGGGCTCTCTCTCGTCTGCTCTGCGCGCGCTG 60  
DB 14 ATGCGTGGCTCCGCGACGAGCGGGCTCTCTCTCGTCTGCTCTGCGCGCGCTG 73  
QY 61 CTGAGGCGGGCTAGGGCTCCCGTGAAGAGCGCGCTCCGCGACCAACGCGCTGGG 120  
DB 74 CTGAGGCGGGCTAGGGCTCCCGTGAAGAGCGCGCTCCGCGACCAACGCGCTGGG 133  
QY 121 AGCTCTCAGAGGCTCGAGAGTCTCAGCTCCCGAGATCTAGGCTCTGAAGAGAG 180  
DB 134 AGCTCTCAGAGGCTCGAGAGTCTCAGCTCCCGAGATCTAGGCTCTGAAGAGAG 193  
QY 181 GAGGAGGACCACTGCTCTCCAGAACCCCTGAGGAGGAGCCACCAACATGATGC 240  
DB 194 GAGGAGGACCACTGCTCTCCAGAACCCCTGAGGAGGAGCCACCAACATGATGC 253  
QY 241 TGGAGTGTACTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
DB 254 TGGAGTGTACTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 313  
QY 301 GAGGTTACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360  
DB 314 GAGGTTACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 373  
QY 361 AGTACCCCTTAACCTGATACCCAGGCTTCAGGCTCCCGAGATCTAGGCTCTGAGGAA 420  
DB 374 AGTACCCCTTAACCTGATACCCAGGCTTCAGGCTCCCGAGATCTAGGCTCTGAGGAA 433  
QY 421 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480  
DB 434 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 493  
QY 481 TGTGGAGTGTACTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
DB 494 TGTGGAGTGTACTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 553  
QY 541 CAGGAGGTTACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600  
DB 554 CAGGAGGTTACTCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 613  
QY 601 TTGAGTACCCCTTAACCTGATACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 639  
DB 614 TTGAGTACCCCTTAACCTGATACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 652

## RESULT 14

BX336623  
LOCUS  
DEFINITION BX336623 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CSODI030Y019 5-PRIME, mRNA sequence.  
BX336623  
ACCESSION BX336623  
971 bp mRNA linear EST 07-APR-2004



```
Qy 181 GAGGAGGACCACTGCTCCCGAGAACCCACCTGAGGAGAGGCCACCAACATGGATGC 240
Db |||||
Qy 195 GAGGAGGACCACTGCTCCCGAGAACCCACCTGAGGAGAGGCCACCAACATGGATGC 254
Db |||||
Qy 241 TGGACTGTCACTGAGCGAGAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 300
Db |||||
Qy 255 TGGACTGTCACTGAGCGAGAGCCATGACCCAGGCAACACACCCCTCCAGGACCCCA 314
Db |||||
Qy 301 GAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATTGGCCAGCACACCTTG 360
Db |||||
Qy 315 GAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATTGGCCAGCACACCTTG 374
Db |||||
Qy 361 AGTACCCCTAACCCCTGATACCCAGGCTTCAGGCTCCCGAGATCCTAGGCTCTGAGGGAA 420
Db |||||
Qy 375 AGTACCCCTAACCCCTGATACCCAGGCTTCAGGCTCCCGAGATCCTAGGCTCTGAGGGAA 434
Db |||||
Qy 421 GAGGAGGAGGACGACTGCTCCCGAGAACCCACCTGCAGGAGAGCTACACCAACATGGA 480
Db |||||
Qy 435 GAGGAGGAGGACGACTGCTCCCGAGAACCCACCTGCAGGAGAGCTACACCAACATGGA 494
Db |||||
Qy 481 TGTGGACTGTCACTGAGCGAGAGCCCTGACCCAGGGAATGCCAGCCTCCAGGACC 540
Db |||||
Qy 495 TGTGGACTGTCACTGAGCGAGAGCCCTGACCCAGGGAATGCCAGCCTCCAGGACC 554
Db |||||
Qy 541 CAGGAGGTTACTCCCTTGTCTGCTGAGCTGCAGAGCTGCAGAAATGGTCCACGCAACC 600
Db |||||
Qy 555 CAGGAGGTTACTCCCTTGTCTGCTGAGCTGCAGAGCTGCAGAAATGGTCCACGCAACC 614
Db |||||
Qy 601 TTGAGTACCCCTAACCCCTGATAACCGGTGACCATCAAG 639
Db |||||
Qy 615 TTGAGTACCCCTAACCCCTGATAACCGGTGACCATCAAG 653
Db |||||
```

Search completed: February 22, 2005, 00:03:54  
Job time : 3102 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 23:41:21 ; Search time 169 Seconds  
(without alignments)  
427.954 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995  
Sequence: 1 LPVKPRLRGPRGSLRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_16Dec04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	995	100.0	571	4	AAB49765 Human pro
2	995	100.0	571	4	AAB88393 Human mem
3	430	43.2	81	4	AAM13752 Peptide #
4	430	43.2	81	4	ABB32685 Peptide #
5	430	43.2	81	4	AAM26150 Peptide #
6	430	43.2	81	4	ABB27526 Human pep
7	430	43.2	81	4	ABB18176 Protein #
8	430	43.2	81	4	AAM65885 Human bon
9	430	43.2	81	4	AAM53507 Human bra
10	430	43.2	81	4	ABG47540 Human liv
11	430	43.2	81	4	AAM01497 Peptide #
12	430	43.2	81	5	ABG35520 Human pep
13	144	14.5	546	4	ABU53252 Human tes
14	144	14.5	1270	8	ADK67911 Human ext
15	144	14.5	1299	4	AAM24322 Human EST
16	144	14.5	1311	8	ADK67912 Human ext
17	144	14.5	1320	7	ADK65819 Angiogene
18	144	14.5	1404	2	AAR26049
19	144	14.5	1404	4	AAB29773 Human meg
20	144	14.5	1404	4	AAB60568 Human meg
21	144	14.5	1404	7	ADK65839 Angiogene
22	144	14.5	1404	8	ADM98014 Human meg
23	144	14.5	1415	4	AAU32262 Novel hum
24	143.5	14.4	513	4	ABU53254 Human tes
25	140.5	14.1	551	4	ABU53253 Human tes

26	127	12.8	399	3	AAB43375	Aab43375 Human ORF
27	122	12.3	1171	4	ABB64387	Abb64387 Drosophil
28	118	11.9	345	8	ABO58798	AbO58798 Human gen
29	115.5	11.6	1518	6	ABJ18375	AbJ18375 Breast ep
30	114.5	11.5	235	6	ADA57458	AdA57458 Human sec
31	114.5	11.5	244	3	AAB38427	Aab38427 Fragment
32	114.5	11.5	288	7	ADD71122	Add71122 Human int
33	114.5	11.5	329	7	ADM04162	Adm04162 Human pro
34	114.5	11.5	406	4	AAU01623	Aau01623 Human sec
35	114.5	11.5	608	8	ADQ66076	Adq66076 Novel hum
36	114.5	11.5	681	4	ABG07912	Abg07912 Novel hum
37	114.5	11.5	726	4	AAB95371	Aab95371 Human pro
38	114.5	11.5	803	4	AAB94325	Aab94325 Human pro
39	114.5	11.5	803	4	AAG67424	Aag67424 Amino aci
40	114.5	11.5	803	5	ABB08164	Abb08164 Human cyt
41	114.5	11.5	803	5	ABG30717	Abg30717 Human Rho
42	114.5	11.5	803	5	ABB98248	Abb98248 Human Rho
43	114.5	11.5	803	8	ABM81720	Abm81720 Tumour-as
44	114.5	11.5	881	4	AAB90819	Aab90819 Human she
45	114.5	11.5	881	4	AAG65820	Aag65820 Human GTP

ALIGNMENTS

RESULT 1  
AAB49765  
ID AAB49765 standard; protein; 571 AA.  
XX  
AC AAB49765;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DB Human proliferation differentiation factor amino acid sequence.  
XX  
KW Human; proliferation differentiation factor; haematopoietic function.  
XX  
OS Homo sapiens.  
XX  
PN W0200104312-A1.  
PD 18-JAN-2001.  
XX  
PF 06-JUL-2000; 2000WO-JP004514.  
XX  
PR 08-JUL-1999; 99JP-00194179.  
PR 18-OCT-1999; 99US-0159586P.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Oca T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;  
XX  
DR WPI; 2001-138354/14.  
XX  
PT N-PSDB; AAF29348.  
XX  
PT Proliferation-differentiation factor protein encoded by PSEC137 cloned from human cDNA library, being hematopoietic factor for inducing differentiation of blood cells, used to maintain hematopoietic function.  
XX  
PS Claim 1; Page 42-44; 49pp; Japanese.  
XX  
CC This invention relates to polynucleotide sequence PSEC137, which encodes a human proliferation differentiation factor protein. Included in the invention is a vector containing the cDNA sequence, a transformant containing the vector, and a process for producing the protein. The protein together with its encoded DNA and antisense DNA are used in drugs to maintain the haematopoietic function. The present sequence represents the human proliferation differentiation factor protein amino acid sequence

SQ Sequence 571 AA;

Query Match 100.0%; Score 995; DB 4; Length 571;

```

Best Local Similarity 100.0%; Pred. No. 1.6e-75;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPVKPRLRGPRGSLTRLAEVASDPRLKEEEAPLLPRTHLQAEPHQHGCWTVTEP 60
Db 27 LPVKKPLRGPRGSLTRLAEVASDPRLKEEEAPLLPRTHLQAEPHQHGCWTVTEP 86
Qy 61 AAMTPGNTTPTPTPEVTPLRLLEQLKPLGLASTTLLSTNPDTQASASDPRLREEEARL 120
Db 87 AAMTPGNTTPTPTPEVTPLRLLEQLKPLGLASTTLLSTNPDTQASASDPRLREEEARL 146
Qy 121 LPRTHLQAEHLQHGCWTVTEPAALTPGNATPPRTQEVTPLELLEQLKPLVHATLSTPNP 180
Db 147 LPRTHLQAEHLQHGCWTVTEPAALTPGNATPPRTQEVTPLELLEQLKPLVHATLSTPNP 206
Qy 181 DNQVTIK 187
Db 207 DNQVTIK 213

RESULT 2
ID AAB88393 standard; protein; 571 AA.
XX AC AAB88393;
XX DT 23-MAY-2001 (first entry)
XX DE Human membrane or secretory protein clone PSEC0137.
XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;
XX KW rheumatoid arthritis; diabetes.
XX OS Homo sapiens.
XX PN EP1067182-A2.
XX PD 10-JAN-2001.
XX PF 07-JUL-2000; 2000EP-00114090.
XX PR 08-JUL-1999; 99JP-00194179.
XX PR 11-JAN-2000; 2000JP-00118775.
XX PR 02-MAY-2000; 2000JP-00183766.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX DR WPI; 2001-093989/11.
XX DR N-PSDB; AAF93820.
XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX PT gene therapy or as candidate target molecules in drug development.
XX PS Claim 1; SEQ ID NO 154; 609pp + Sequence Listing; English.
XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX CC which encode human secretory or membrane proteins represented by AAB88317
XX CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
XX CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
XX CC invention. The invention also includes methods for the production of
XX CC antibodies directed against the proteins, and cDNA sequences, which can
XX CC be used in vaccines. The polynucleotide sequences can be used in gene
XX CC therapy. The polynucleotide sequences and the proteins they encode may be
XX CC used in the prevention, treatment and diagnosis of diseases associated
XX CC with inappropriate secretory protein/membrane protein expression. The
XX CC nucleic acids and complementary sequences may also be used as DNA probes
XX CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
XX CC and quantitate the presence of similar nucleic acid sequences in samples.
XX CC They may also be used to study the expression and function of secretory
XX CC proteins/membrane polypeptides and their role in metabolism. The
XX CC polypeptides may be used as antigens in the production of antibodies

```

CC against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbant assay (ELISA). Examples of diseases which may be CC treated include rheumatoid arthritis and diabetes

CC Sequence 571 AA;

Query Match 100.0%; Score 995; DB 4; Length 571;  
Best Local Similarity 100.0%; Pred. No. 1.6e-75;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPVKKPLRGPRGSLTRLAEVASDPRLKEEEAPLLPRTHLQAEPHQHGCWTVTEP 60  
Db 27 LPVKKPLRGPRGSLTRLAEVASDPRLKEEEAPLLPRTHLQAEPHQHGCWTVTEP 86  
Qy 61 AAMTPGNTTPTPTPEVTPLRLLEQLKPLGLASTTLLSTNPDTQASASDPRLREEEARL 120  
Db 87 AAMTPGNTTPTPTPEVTPLRLLEQLKPLGLASTTLLSTNPDTQASASDPRLREEEARL 146  
Qy 121 LPRTHLQAEHLQHGCWTVTEPAALTPGNATPPRTQEVTPLELLEQLKPLVHATLSTPNP 180  
Db 147 LPRTHLQAEHLQHGCWTVTEPAALTPGNATPPRTQEVTPLELLEQLKPLVHATLSTPNP 206  
Qy 181 DNQVTIK 187  
Db 207 DNQVTIK 213

RESULT 3  
AAM13752  
ID AAM13752 standard; protein; 81 AA.  
XX AC AAM13752;  
XX DT 12-OCT-2001 (first entry)  
XX DE Peptide #186 encoded by probe for measuring cervical gene expression.  
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;  
XX KW cervical cancer.  
XX OS Homo sapiens.  
XX PN WO200157278-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US0000670.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR WPI; 2001-488901/53.  
XX CC Human genome-derived single exon nucleic acid probes useful for analyzing  
XX CC gene expression in human cervical epithelial cells.  
XX PS Claim 27; SEQ ID NO 18578; 487pp; English.  
XX CC The present invention relates to human single exon nucleic acid probes  
XX CC (SENP; see AAI10068-AA128459). The present sequence is a peptide encoded

CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 81 AA;

Query Match 43.2%; Score 430; DB 4; Length 81;  
 Best Local Similarity 97.5%; Pred. No. 8.3e-29;  
 Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 22 VSASPDPRPLKEEEAELLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81  
 ID AAM26150 standard; protein; 81 AA.  
 DB 1 VSASPDPRPLKEEEAELLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60

QY 82 ELQKLPLGLASTLSTPNPDQ 102  
 ID ELQKLPLGLANTLSTPNPDQ 81

RESULT 4  
 ID ABB32685 standard; peptide; 81 AA.  
 AC ABB32685;

DT 04-FEB-2002 (first entry)

DE Peptide #191 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human fetal liver.

XX Claim 27; SEQ ID NO 25320; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

XX human gene expression in a sample derived from human foetal liver. The

XX single exon nucleic acid probes may be used for predicting, measuring and

XX displaying gene expression in samples derived from human fetal liver. The

XX present sequence is a peptide encoded by a single exon nucleic acid probe

XX of the invention. Note: The sequence data for this patent did not form

XX part of the printed specification, but was obtained in electronic format

XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 81 AA;

Query Match 43.2%; Score 430; DB 4; Length 81;  
 Best Local Similarity 97.5%; Pred. No. 8.3e-29;  
 Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 22 VSASPDPRPLKEEEAELLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81  
 ID AAM26150 standard; protein; 81 AA.  
 DB 1 VSASPDPRPLKEEEAELLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60

QY 82 ELQKLPLGLASTLSTPNPDQ 102  
 ID ELQKLPLGLANTLSTPNPDQ 81

RESULT 5  
 ID AAM26150 standard; protein; 81 AA.  
 AC AAM26150;

XX 17-OCT-2001 (first entry)

XX Peptide #187 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human placenta.

XX Claim 27; SEQ ID NO 26419; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs:

XX see AA131315-AA157546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders

XX Sequence 81 AA;

Query Match 43.2%; Score 430; DB 4; Length 81;  
 Best Local Similarity 97.5%; Pred. No. 8.3e-29;  
 Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 22 VSASPDPRPLKEEEAELLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81  
 DB 1 VSASPDPRPLKEEEAELLPRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60

QY 82 ELQKLPLGLASTLSTPNPDQ 102  
 ID ELQKLPLGLANTLSTPNPDQ 81

```
Db      61 ELQKLPGLANTTLSTPNPDQ 81
RESULT 6
ABB27526
ID      ABB27526 standard; peptide; 81 AA.
XX
AC      ABB27526;
XX
DT      01-FEB-2002 (first entry)
XX
DE      Human peptide #177 encoded by breast cell single exon nucleic acid probe.
XX
KW      Human; microarray; single exon probe; gene expression; breast; disease;
XX      cancer.
XX
OS      Homo sapiens.
XX
PN      WO200157271-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000662.
XX
PR      04-FEB-2000; 2000US-0180312P.
XX      26-MAY-2000; 2000US-0207456P.
XX      30-JUN-2000; 2000US-00608408.
XX      03-AUG-2000; 2000US-00632366.
XX      21-SEP-2000; 2000US-0234687P.
XX      27-SEP-2000; 2000US-0236359P.
XX      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-496933/54.
XX
PT      New spatially-addressable set of single exon nucleic acid probes, useful
XX      for measuring gene expression in sample derived from human breast,
XX      comprises number of single exon nucleic acid probes.
XX
PS      Claim 27; SEQ ID NO 10494; 327pp + Sequence Listing; English.
XX
CC      The invention relates to a spatially-addressable set of single exon
XX      nucleic acid probes for measuring gene expression in a sample derived
XX      from human breast and BT 474 cells. The method involves contacting the
XX      probes with a collection of detectably labelled nucleic acids derived
XX      from mRNA of human breast, and then measuring the label bound to each
XX      probe of the microarray. The probes are useful for verifying the
XX      expression of regions of genomic DNA predicted to encode proteins. They
XX      are useful for gene discovery, and for determining predisposition and/or
XX      prognosing breast disease. Gene expression analysis is useful for
XX      assessing the toxicity of chemical agents on cells. The microarray of
XX      this invention presents a far greater diversity of probes for measuring
XX      gene expression, with far less bias than expressed sequence tag
XX      microarrays. The method is suitable for rapid production of functional
XX      information from genomic sequence. The present sequence is a peptide
XX      encoded by a single exon nucleic acid probe of the invention. Note: The
XX      sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 81 AA;
Query Match      43.2%; Score 430; DB 4; Length 81;
Best Local Similarity 97.5%; Pred. NO. 8.3e-29;
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      22 VSASDPRLKKEEAPLLPRTHLQAEPHQGCWTVTEPAAMTGCNTTTPRTPEVTPLRL 81
DB      1 VSASDPRLKKEEAPLLPRTHLQAEPHQGCWTVTEPAAMTGCNTTTPRTPEVTPLRL 60
QY      82 ELQKLPGLASTTLSTPNPDQ 102
DB      61 ELQKLPGLANTTLSTPNPDQ 81
```

```
QY      82 ELQKLPGLASTTLSTPNPDQ 102
DB      61 ELQKLPGLANTTLSTPNPDQ 81
RESULT 7
ABB18176
ID      ABB18176 standard; protein; 81 AA.
XX
AC      ABB18176;
XX
DT      23-JAN-2002 (first entry)
XX
DE      Protein #175 encoded by probe for measuring heart cell gene expression.
XX
KW      Human; gene expression; heart; microarray; vascular system;
XX      cardiovascular disease; hypertension; cardiac arrhythmia;
XX      congenital heart disease.
XX
OS      Homo sapiens.
XX
PN      WO200157274-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000666.
XX
PR      04-FEB-2000; 2000US-0180312P.
XX      26-MAY-2000; 2000US-0207456P.
XX      30-JUN-2000; 2000US-00608408.
XX      03-AUG-2000; 2000US-00632366.
XX      21-SEP-2000; 2000US-0234687P.
XX      27-SEP-2000; 2000US-0236359P.
XX      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488899/53.
XX
PT      Single exon nucleic acid probes for analyzing gene expression in human
XX      hearts.
XX
PS      Claim 15; SEQ ID NO 19946; 530pp; English.
XX
CC      The present invention relates to single exon nucleic acid probes for
XX      measuring human gene expression in a sample derived from human heart (see
XX      ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX      probe. The probes may be used for predicting, measuring and displaying
XX      gene expression in samples derived from the human heart via microarrays.
XX      By measuring gene expression, the probes are useful for predicting,
XX      diagnosing, grading, staging, monitoring and prognosing diseases of the
XX      human heart and vascular system e.g. cardiovascular disease.
XX      hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX      sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 81 AA;
Query Match      43.2%; Score 430; DB 4; Length 81;
Best Local Similarity 97.5%; Pred. NO. 8.3e-29;
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      22 VSASDPRLKKEEAPLLPRTHLQAEPHQGCWTVTEPAAMTGCNTTTPRTPEVTPLRL 81
DB      1 VSASDPRLKKEEAPLLPRTHLQAEPHQGCWTVTEPAAMTGCNTTTPRTPEVTPLRL 60
QY      82 ELQKLPGLASTTLSTPNPDQ 102
DB      61 ELQKLPGLANTTLSTPNPDQ 81
```

```
RESULT 8
AAM65885
ID AAM65885 standard; protein; 81 AA.
XX AC AAM65885;
XX AC
XX AC
DT 06-NOV-2001 (first entry)
XX XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 26191.
XX XX
XX Human: bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX KW
XX OS Homo sapiens.
XX XX
XX PN W0200157276-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX 30-JAN-2001; 2001WO-US000668.
XX XX
XX 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI
XX WPI; 2001-488900/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human bone marrow.
XX PT
XX PS Example 4; SEQ ID NO 26191; 658pp + Sequence Listing; English.
XX XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention
XX XX
XX SQ Sequence 81 AA;
XX XX
XX Query Match 43.2%; Score 430; DB 4; Length 81;
XX Best Local Similarity 97.5%; Pred. No. 8.3e-29;
XX Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
Qy 22 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNATPPRTPEVTPLRL 81
Db 1 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNATPPRTPEVTPLRL 60
Qy 82 ELQKLPGLASTTLSTPNPDQ 102
Db 61 ELQKLPGLASTTLSTPNPDQ 81
XX
RESULT 9
AAM53507
ID AAM53507 standard; protein; 81 AA.
XX AC AAM53507;
XX AC
XX AC
DT 05-NOV-2001 (first entry)
XX XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 25612.
XX XX
XX XX
```

```
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX PN W0200157275-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX 30-JAN-2001; 2001WO-US000667.
XX XX
XX 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI
XX WPI; 2001-483446/52.
XX DR
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains.
XX PT
XX PS Example 4; SEQ ID NO 25612; 650pp + Sequence Listing; English.
XX XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention
XX XX
XX SQ Sequence 81 AA;
XX XX
XX Query Match 43.2%; Score 430; DB 4; Length 81;
XX Best Local Similarity 97.5%; Pred. No. 8.3e-29;
XX Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
Qy 22 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNATPPRTPEVTPLRL 81
Db 1 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNATPPRTPEVTPLRL 60
Qy 82 ELQKLPGLASTTLSTPNPDQ 102
Db 61 ELQKLPGLASTTLSTPNPDQ 81
XX
RESULT 10
ABG47540
ID ABG47540 standard; peptide; 81 AA.
XX AC ABG47540;
XX AC
XX AC
DT 25-FEB-2003 (first entry)
XX XX
XX Human liver peptide, SEQ ID No 26188.
XX XX
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX KW
XX OS Homo sapiens.
XX XX
XX PN W0200157273-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX 30-JAN-2001; 2001WO-US000664.
```

```

XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488898/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX Claim 27; SEQ ID NO 26188; 658pp; English.
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinemia, hyperlipidemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 81 AA;
XX
Query Match 43.2%; Score 430; DB 4; Length 81;
Best Local Similarity 97.5%; Pred. No. 8.3e-29;
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 22 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81
DB 1 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60
QY 82 ELQKPLGLASTTLSTPNPDQ 102
DB 61 ELQKPLGLANTTLSTPNPDQ 81
RESULT 11
AA001497
ID AA001497 standard; protein; 81 AA.
XX AA001497;
XX
XX 09-OCT-2001 (first entry)
XX Peptide #179 encoded by probe for measuring human breast gene expression.
XX Probe; human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX Homo sapiens.
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US0000661.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.

```

```

PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.
XX Claim 27; SEQ ID NO 10237; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes
XX (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative,
XX breast disease and non-carcinoma tumours. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 81 AA;
XX
Query Match 43.2%; Score 430; DB 4; Length 81;
Best Local Similarity 97.5%; Pred. No. 8.3e-29;
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 22 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 81
DB 1 VSASDPRLKKEEAPLLRTHLQAEPHQHCWTVTEPAAMTGNTPPTPEVTPLRL 60
QY 82 ELQKPLGLASTTLSTPNPDQ 102
DB 61 ELQKPLGLANTTLSTPNPDQ 81
RESULT 12
ABG35520
ID ABG35520 standard; peptide; 81 AA.
XX AC ABG35520;
XX
XX 19-AUG-2002 (first entry)
XX Human peptide encoded by genome-derived single exon probe SEQ ID 25185.
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX Homo sapiens.
XX OS
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US0000665.

```

XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
XX Claim 27; SEQ ID NO 25185; 634pp; English.  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridize at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with hybridisation to a single exon  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 81 AA;  
XX Query Match 43.2%; Score 430; DB 5; Length 81;  
XX Best Local Similarity 97.5%; Pred. No. 8.3e-29;  
XX Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 22 VSASDPRLKEEAEALLRTHLQAEPHQHGCVTTEPAAMTGNTPRTTEVTPLRL 81  
DB 1 VSASDPRLKEEAEALLRTHLQAEPHQHGCVTTEPAAMTGNTPRTTEVTPLRL 60  
QY 82 ELOKPLGLASTLSTNPDTQ 102  
DB 61 ELOKPLGLASTLSTNPDTQ 81

## RESULT 13

ABUS3252

XX ABUS3252 standard; protein; 546 AA.  
AC ABUS3252;  
XX 14-APR-2003 (first entry)  
XX Human testes-derived DKFZphtes3\_4019 homologue #1.  
DE Human; gene therapy; vaccine; disease treatment; detection.  
KW Homo sapiens.  
OS WO200112659-A2.  
XX PN 22-FEB-2001.  
XX PD 18-AUG-2000; 2000WO-IB001496.  
XX PF 18-AUG-1999; 99US-0149499P.  
XX PR 28-SEP-1999; 99US-0156503P.  
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX WIemann S;  
XX WPI; 2001-327840/34.  
XX Nucleic acids having the sequences of clones isolated from libraries of  
PT different human tissues, useful in recombinant DNA methodologies.  
XX Example III; Page 892; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention  
XX Sequence 546 AA;

Query Match 14.5%; Score 144; DB 4; Length 546;  
Best Local Similarity 28.4%; Pred. No. 0.001;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASFDPRLKEEAEALLRTHLQAEPHQHGCVTTEPAAMT-----PGNTTPP 71  
DB 68 TTTPKEPAPTPKPEAPTPPKKAPAPTPPKPEAPTPPKKAPAPTPPKKAPAPTPPK 127  
QY 72 RTPEVTPLRLLEQLKPLGLASTLSTNPDTQASASP-----DPRPLREEEAEALLRTHLQ 127  
DB 128 ETAPTPKLTPTTPEKLAFTTPEKAPTPPEELAPTTPEPTTPEEAPTPPKAAP 187  
QY 128 AELHQHGCVTTEPAALT---PGNATPPRTOEVTPLLELOKLPELVHATLSTNP 180  
DB 188 NTPKEPAPTPKPEAPTPPKPEAPTPPKETAPTPKGTAPTKLKEPAPTPPKKAP 243

## RESULT 14

ADK67911

XX ADK67911 standard; protein; 1270 AA.  
XX AC ADK67911;  
XX DT 06-MAY-2004 (first entry)  
XX

Human extracellular messenger (EXMES) polypeptide.

Human; extracellular messenger; EXMES; respiratory-gen.; antiallergic; antiasthmatic; antiinflammatory; antidiabetic; neuroprotective; muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiporiatic; virucide; fungicide; antiparasitic; protozoacide; antihelminthic; cytostatic; gene therapy.

Homo sapiens.

Key Location/Qualifiers  
Peptide 1..29  
/label= Signal\_peptide  
/note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29 according to identification method"

WO2004013292-A2.

12-FEB-2004.

30-JUL-2003; 2003WO-US024084.

02-AUG-2002; 2002US-0400810P.

19-SEP-2002; 2002US-0412197P.

04-OCT-2002; 2002US-0416004P.

08-NOV-2002; 2002US-0424862P.

(INCY-) INCYTE CORP.

Elliot VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;

Lee SY;

WPI; 2004-157116/15.

N-PSDB; ADK67916.

New extracellular messengers and nucleic acids, useful for diagnosing, treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or autoimmune thyroiditis.

Claim 59; SEQ ID NO 4; 165pp; English.

The present sequence is that of novel human extracellular messenger (EXMES) incyte ID NO: 7513017CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES polynucleotides and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES, especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or their encoding polynucleotides for facilitating the drug discovery process, including determining of efficacy, dosage, toxicity and pharmacology, and for investigating the pathogenesis of diseases and medical conditions.

Sequence 1270 AA;  
Query Match 14.5%; Score 144; DB 8; Length 1270;  
Best Local Similarity 28.4%; Pred. No. 0.0026;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

23 SASPDPRPLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
421 TTTKEPAPTTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEK 480

72 RTPVTPRLRLQKLPGLASTTLPNDTQASASP-----DPRPLREERARLLPRTHLQ 127

128 AELHQHCWTVTEPAALT---PGNATPPRTQVTPVTLLELQKLPELVHATLSTNP 180

Db 481 ETAPTTPKLTPTTPEKLAFTTPEKPAFTTPEELAPTTPEPTTPEEPAPTTPKAAAP 540  
Qy 128 AELHQHCWTVTEPAALT---PGNATPPRTQVTPVTLLELQKLPELVHATLSTNP 180  
Db 541 NTPKEPAPTTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEK 596

RESULT 15

AA24322

ID AA24322 standard; protein; 1299 AA.

XX AC AA24322;

XX DT 12-OCT-2001 (first entry)

XX DE Human EST encoded protein SEQ ID NO: 1847.

XX DE Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

XX KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

XX KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;

XX KW gene therapy; nutrition.

XX OS Homo sapiens.

XX PN WO200154477-A2.

XX PD 02-AUG-2001.

XX PF 25-JAN-2001; 2001WO-US002687.

XX PR 25-JAN-2000; 2000US-00491404.

XX PR 17-JUL-2000; 2000US-00617746.

XX PR 03-AUG-2000; 2000US-00631451.

XX PR 15-SEP-2000; 2000US-00663870.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX DR WPI; 2001-476164/51.

XX DR N-PSDB; AAH98981.

XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising antibodies and research use.  
XX PS Claim 20; Page 1198-1201; 1275pp; English.  
XX CC The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a protein of the invention

XX SQ Sequence 1299 AA;

Query Match 14.5%; Score 144; DB 4; Length 1299;  
Best Local Similarity 28.4%; Pred. No. 0.0027;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASPDPRPLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71

Db 555 TTTKEPAPTTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEKPAFTTPEK 614  
Qy 72 RTPVTPRLRLQKLPGLASTTLPNDTQASASP-----DPRPLREERARLLPRTHLQ 127

Db 615 ETAPTTPKLTPTTPEKLAFTTPEKPAFTTPEELAPTTPEPTTPEEPAPTTPKAAAP 674  
Qy 128 AELHQHCWTVTEPAALT---PGNATPPRTQVTPVTLLELQKLPELVHATLSTNP 180



Db 675 NTPKEPAPPTPKKEPAPPTPKETAPPTPKGTAPPTLKEPAPPTPKKEPAP 730

Search completed: February 18, 2005, 23:56:01  
Job time : 172 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005. Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 23:50:51 ; Search time 43 Seconds  
(without alignments)  
324.637 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKPLRGRPGSLRRLA.....PELVHATLSTPNPDNQVTK 187

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	14.5	941	4	US-07-757-022B-14
2	144	14.5	1022	4	US-07-757-022B-84
3	144	14.5	1038	4	US-07-757-022B-74
4	144	14.5	1049	4	US-07-757-022B-58
5	144	14.5	1140	4	US-07-757-022B-104
6	144	14.5	1270	4	US-07-757-022B-44
7	144	14.5	1311	4	US-07-757-022B-42
8	144	14.5	1313	4	US-07-757-022B-142
9	144	14.5	1314	4	US-07-757-022B-50
10	144	14.5	1320	4	US-07-757-022B-46
11	144	14.5	1320	4	US-07-757-022B-60
12	144	14.5	1320	4	US-10-164-595-58
13	144	14.5	1354	4	US-07-757-022B-48
14	144	14.5	1361	4	US-07-757-022B-40
15	144	14.5	1363	4	US-07-757-022B-52
16	144	14.5	1404	4	US-07-757-022B-2
17	144	14.5	1404	4	US-07-757-022B-62
18	144	14.5	1404	4	US-10-164-595-78
19	144	14.5	1404	4	US-09-298-970A-1
20	144	14.5	1411	4	US-09-949-016-10827
21	113	11.4	5179	4	US-09-538-092-1258
22	112	11.3	183	6	5168049-3
23	112	11.3	183	6	5168049-3
24	111.5	11.2	267	3	US-08-818-112-142
25	111.5	11.2	267	3	US-08-818-111-137
26	111.5	11.2	267	3	US-09-056-556-142
27	111.5	11.2	267	4	US-09-072-596-137

Sequence 142, App  
Sequence 945, App  
Sequence 17307, A  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 9700, Ap  
Sequence 8969, Ap  
Sequence 736, App  
Sequence 6668, Ap  
Sequence 2, Appli  
Sequence 143, App  
Sequence 143, App  
Sequence 34, Appl  
Sequence 34, Appl  
Sequence 110, App  
Sequence 4, Appli  
Sequence 90, Appl  
Sequence 90, Appl  
Sequence 5258, Ap

28 111.5 11.2 267 4 US-09-072-967-142  
29 108.5 10.9 705 4 US-09-538-092-945  
30 108.5 10.9 827 4 US-09-248-796A-17307  
31 107.5 10.8 696 3 US-08-906-865-4  
32 107.5 10.8 696 4 US-09-129-668-4  
33 107.5 10.8 713 4 US-09-949-016-9700  
34 106 10.7 288 4 US-09-949-016-8969  
35 105.5 10.6 1530 4 US-09-976-594-736  
36 105.5 10.6 1530 4 US-09-949-016-6668  
37 105 10.6 486 1 US-08-450-360-2  
38 105 10.6 699 4 US-10-237-551-143  
39 105 10.6 699 4 US-10-237-551-254  
40 105 10.6 1215 4 US-10-029-180-34  
41 105 10.6 1244 4 US-10-029-180-110  
42 104.5 10.5 447 1 US-08-450-360-4  
43 103.5 10.4 258 3 US-08-961-083-90  
44 103.5 10.4 258 4 US-09-536-784-90  
45 103.5 10.4 305 4 US-09-107-532A-5258

#### ALIGNMENTS

#### RESULT 1

US-07-757-022B-14  
; Sequence 14, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Geener, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 941 amino acids  
; TYPE: AMINO ACID

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-14

Query Match      14.5%; Score 144; DB 4; Length 941;
Best Local Similarity 28.4%; Pred. No. 2.4e-05;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 356 TTTKEPATTTPKEPATTTPKPAATTPKEPATTTPKPAATTPKPAATTPKPAATTPK 415
QY 72 RTPEVTPRLRLLEQKLPGLASTTSTPNPDTOASASP-----DPRPLREEEERALLPRTHLQ 127
Db 416 ETAPTTPKLTPPTPEKLAATTPKEPATTTPKEPATTTPPEELAPTTPEEPTTTTPEEPATTPKAAP 475
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTOEVTPLLELQKLPELVHATLSTNP 180
Db 476 NTPKEPATTTPKEPATTTPKPAATTPKETAATTPKGTAPTTLKEPATTTPKPAAP 531

RESULT 2
US-07-757-022B-84
; Sequence 84, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-74

Query Match      14.5%; Score 144; DB 4; Length 1022;
Best Local Similarity 28.4%; Pred. No. 2.7e-05;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 437 TTTKEPATTTPKEPATTTPKPAATTPKEPATTTPKPAATTPKPAATTPKPAATTPK 496
QY 72 RTPEVTPRLRLLEQKLPGLASTTSTPNPDTOASASP-----DPRPLREEEERALLPRTHLQ 127
Db 497 ETAPTTPKLTPPTPEKLAATTPKEPATTTPKEPATTTPPEELAPTTPEEPTTTTPEEPATTPKAAP 556
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTOEVTPLLELQKLPELVHATLSTNP 180
Db 557 NTPKEPATTTPKEPATTTPKPAATTPKETAATTPKGTAPTTLKEPATTTPKPAAP 612

RESULT 3
US-07-757-022B-74
; Sequence 74, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1038 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-74

```

Query Match 14.5%; Score 144; DB 4; Length 1038;  
Best Local Similarity 28.4%; Pred. No. 2.8e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
DB 421 TTTKEPAPTTPKKPAPTTPKKPAPTTPKKPAPTTTKKPAPTAKEPAPTPK 480  
QY 72 RTPEVTPLRLQLKPLGLASTTLPNDTQASAP-----DPRPLREEEEARLLPRTHLQ 127  
DB 481 ETAPTTPKLTPPTPEKLPAPTTPEKLPAPTTPEELAPTTPEEPTTTPPEPAPTPKAAAP 540  
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTOEVTPLLELQKLDELVHATLSTNP 180  
DB 541 NTPKEPAPTTPKKPAPTTPKKPAPTTPKKPAPTTPKKPAPTTPKKPAP 596

RESULT 4  
US-07-757-022B-58  
; Sequence 58, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1049 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-58

Query Match 14.5%; Score 144; DB 4; Length 1049;  
Best Local Similarity 28.4%; Pred. No. 2.8e-05;

Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
DB 464 TTTKEPAPTTPKKPAPTTPKKPAPTTPKKPAPTTTKKPAPTAKEPAPTPK 523  
QY 72 RTPEVTPLRLQLKPLGLASTTLPNDTQASAP-----DPRPLREEEEARLLPRTHLQ 127  
DB 524 ETAPTTPKLTPPTPEKLPAPTTPEKLPAPTTPEELAPTTPEEPTTTPPEPAPTPKAAAP 583  
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTOEVTPLLELQKLDELVHATLSTNP 180  
DB 584 NTPKEPAPTTPKKPAPTTPKKPAPTTPKKPAPTTPKKPAPTTPKKPAP 639

RESULT 5  
US-07-757-022B-104  
; Sequence 104, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 104:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1140 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-104

Query Match 14.5%; Score 144; DB 4; Length 1140;  
Best Local Similarity 28.4%; Pred. No. 3.1e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

[illegible]

RESULT 6  
US-07-757-022B-44  
; Sequence 44, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1270 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-44

Query Match 14.5%; Score 144; DB 4; Length 1270;  
Best Local Similarity 28.4%; Pred. No. 3.6e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASPDRLKKEEAPLLPETHLQAEPHQGCWVTPEAAWT-----PGNTTTP 71  
          :::               :::

421	TTTKEPAPTTTPKEPAPTTTPKPAEPTTPKEPAPTTTPKEPAPTTTKKPAETAPKEPAPTTTPK	Db
72	RTPEVTPRLLELOKLPGLASTTSLTSPNDTOASASP-----DPRPLRSEEEAKLLPRTHLQ	Qy
481	ETAPTTTCKLUTTPKEAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKAAAP	Db
128	AEHQHCQVTEPAALT-----PGNATPPRTQEVTPLLLELOKLPELVHATLSTNP	Qy
541	NTPEKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKPKAP	Db

```

RESULT 7
US-07-757-022B-42
; Sequence 42, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA: US 07/390,901
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO.: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-42

```

```

Query Match      14.5%; Score 144; DB 4; Length 1311;
Best Local Similarity 28.4%; Pred.No. 3.7e-05;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY   23 SASPDPRPLKBEAEALLPRTHLQAEPHQHCQWTVTEPAAMT-----PGNTTPP 71
      ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db    462 TTTTKKPAPTTTKPPAPTTKKPAPTPTTKPEAPITTKPEAPITTTKKPAPTAKGEPAPITPK 521

```

Db	524	ETATPTTKLTTTPEKLA	PTTPEKAP	PTTPEELAP	TTPEBPT	TTPEBAP	TTPKAAAP	583
Qy	128	AELHQHGCTWTBPAAL	---	PGNAP	PTQEV	PTLLEL	QKLPELV	ATLSTPNP 180
Db	584	NTPKAPATPTTKGAP	PTTTPKEAP	TTTPKE	ATPTPK	GTAP	TTLKE	PAP 639

```

RESULT 9
US-07-757-022B-50
; Sequence 50, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910

CLASSIFICATION: 530  
PRIOR APPLICATION DATA: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:

```

/ NAME: Caerr, Luann
/
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)876-1170
/ TELEFAX: (617)876-5851
/ INFORMATION FOR SEQ ID NO: 50:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1314 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-07-757-022B-50

```

Query Match 14.5%; Score 144; DB 4; Length 1314;  
Best Local Similarity 28.4%; Pred. No. 3.7e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 1

QY 23 SASPPRPLKEEEEAPLLPRTHLQAEPHQHGCTVTEPAAMT-----PGNTTPP 71

Qy	72	RTPEVTPRLRELOKLPGLASTLTSTPNPDQTASASP-----DPRPLREEEERALLPPTHLQ	127
Db	525	ETAPATPKLLPTTPEKLAPTTPPEKPAPTTPPEELAPTTPEEPPTTPEEPAPTTPPEAPATTPEAPAAAP	584

QY 128 AELHQGCVTTEPAALT---PGNATPPRTQEVTPLELLELQKLPVLHATLSTNP 180  
Db 585 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEPATTTPKKAP 640

RESULT 10  
US-07-757-022B-46  
; Sequence 46, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757.022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-46

Query Match 14.5%; Score 144; DB 4; Length 1320;  
Best Local Similarity 28.4%; Pred. No. 3.7e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;  
QY 23 SASPPRLKEEERAPLLPRTHLQAEPHQHCWVTEPAAMT-----PGNTTTP 71  
Db 471 TTTKEPATTTPKEPATTTPKEPATTTPKEPATTTPKAPAPAPKEPATTTPK 530  
QY 72 RPEVTPRLLELQKLPGLASTLTSTPNPDQASAP-----DPRPLREEEERALLPRTHLQ 127  
Db 531 ETAPTTPKLTTPTEKLAFTTPEKAPATTPELAFTTPEEFTTTPPEPATTTPKAAAP 590  
QY 128 AELHQGCVTTEPAALT---PGNATPPRTQEVTPLELLELQKLPVLHATLSTNP 180

Db 591 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEPATTTPKKAP 646  
RESULT 11  
US-07-757-022B-60  
; Sequence 60, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757.022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 60:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-60

Query Match 14.5%; Score 144; DB 4; Length 1320;  
Best Local Similarity 28.4%; Pred. No. 3.7e-05;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;  
QY 23 SASPPRLKEEERAPLLPRTHLQAEPHQHCWVTEPAAMT-----PGNTTTP 71  
Db 471 TTTKEPATTTPKEPATTTPKEPATTTPKEPATTTPKAPAPAPKEPATTTPK 530  
QY 72 RPEVTPRLLELQKLPGLASTLTSTPNPDQASAP-----DPRPLREEEERALLPRTHLQ 127  
Db 531 ETAPTTPKLTTPTEKLAFTTPEKAPATTPELAFTTPEEFTTTPPEPATTTPKAAAP 590  
QY 128 AELHQGCVTTEPAALT---PGNATPPRTQEVTPLELLELQKLPVLHATLSTNP 180  
Db 591 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEPATTTPKKAP 646





```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-40

Query Match 14.5%; Score 144; DB 4; Length 1361;
Best Local Similarity 28.4%; Pred. No. 3.9e-05;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLKKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 512 TTTKEPATTTPKEPATTTPKPAATTPKEPATTTPKPAATTPKPAATTPKPAATTPK 571
QY 72 RTPEVTPRLRLQKLPGLASTTLPNDPTQASAP-----DPRPLREBEEERALLPRTHLQ 127
Db 572 ETAPTTPKKLTPPTPEKLAATTPKEPATTTPPEELAPTTPEBPTPTTPEEPATTPKAAAP 631
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTQEVTPLLLELQKLPVLVHATLSTNP 180
Db 632 NTPKEPATTTPKEPATTTPKEPATTTPKETATPTTKGTAPTTLKEPATTTPKKPAP 687

;
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-52

Query Match 14.5%; Score 144; DB 4; Length 1363;
Best Local Similarity 28.4%; Pred. No. 3.9e-05;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

QY 23 SASDPRLKKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 514 TTTKEPATTTPKEPATTTPKPAATTPKEPATTTPKPAATTPKPAATTPKPAATTPK 573
QY 72 RTPEVTPRLRLQKLPGLASTTLPNDPTQASAP-----DPRPLREBEEERALLPRTHLQ 127
Db 574 ETAPTTPKKLTPPTPEKLAATTPKEPATTTPPEELAPTTPEBPTPTTPEEPATTPKAAAP 633
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTQEVTPLLLELQKLPVLVHATLSTNP 180
Db 634 NTPKEPATTTPKEPATTTPKEPATTTPKETATPTTKGTAPTTLKEPATTTPKKPAP 689

Search completed: February 19, 2005, 00:00:40
Job time : 48 secs
```

```
RESULT 15
US-07-757-022B-52
; Sequence 52, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2005, 23:59:14 ; Search time 130 Seconds  
(without alignments)

470.724 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKKPLRGRPGSLRLA.....PELVHATLSTPNPDNQVTK 187

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB ID	Description
1	933.5	93.8	658	11	US-09-939-853A-94
2	879	88.3	538	15	US-10-074-978A-403
3	430	43.2	81	9	US-09-864-761-33474
4	144	14.5	941	13	US-10-124-557-14
5	144	14.5	1022	13	US-10-124-557-84
6	144	14.5	1038	13	US-10-124-557-74
7	144	14.5	1049	13	US-10-124-557-58
8	144	14.5	1140	13	US-10-124-557-104
9	144	14.5	1270	13	US-10-124-557-44
10	144	14.5	1311	13	US-10-124-557-42
11	144	14.5	1313	13	US-10-124-557-142
12	144	14.5	1314	13	US-10-124-557-50
13	144	14.5	1320	13	US-10-124-557-46

14	144	14.5	1320	13	US-10-124-557-60	Sequence 60, Appl
15	144	14.5	1354	13	US-10-124-557-48	Sequence 48, Appl
16	144	14.5	1361	13	US-10-124-557-40	Sequence 40, Appl
17	144	14.5	1363	13	US-10-124-557-52	Sequence 52, Appl
18	144	14.5	1404	9	US-09-802-207-30	Sequence 30, Appl
19	144	14.5	1404	11	US-09-897-188-1	Sequence 1, Appl
20	144	14.5	1404	13	US-10-124-557-2	Sequence 2, Appl
21	144	14.5	1404	13	US-10-124-557-62	Sequence 62, Appl
22	144	14.5	1404	17	US-10-868-577A-55	Sequence 55, Appl
23	139.5	13.0	404	16	US-10-437-963-116565	Sequence 116565,
24	134.5	12.5	410	16	US-10-437-963-196665	Sequence 196665,
25	132.5	12.3	466	16	US-10-437-963-167964	Sequence 167964,
26	119	12.0	481	16	US-10-437-963-176449	Sequence 176449,
27	118.5	11.9	468	16	US-10-437-963-181077	Sequence 181077,
28	118	11.9	345	14	US-10-029-386-32432	Sequence 32432, A
29	117.5	11.8	442	16	US-10-437-963-197867	Sequence 197867,
30	117.5	11.8	537	16	US-10-437-963-188969	Sequence 188969,
31	116.5	11.7	19723	15	US-10-084-846A-5	Sequence 5, Appl
32	115.5	11.6	1518	11	US-09-989-890-184	Sequence 184, App
33	115	11.6	582	14	US-10-156-761-11938	Sequence 11938, A
34	114.5	11.5	235	14	US-10-050-704-102	Sequence 102, App
35	114.5	11.5	235	16	US-10-798-512-102	Sequence 102, App
36	114.5	11.5	244	14	US-10-050-704-209	Sequence 209, App
37	114.5	11.5	244	16	US-10-798-512-209	Sequence 209, App
38	114.5	11.5	329	15	US-10-108-260A-2847	Sequence 2847, App
39	114.5	11.5	726	9	US-09-770-689A-4	Sequence 4, Appl
40	114.5	11.5	803	9	US-09-770-689A-2	Sequence 2, Appl
41	114.5	11.5	803	14	US-10-059-585-10	Sequence 10, Appl
42	114.5	11.5	803	15	US-10-415-187-8	Sequence 8, Appl
43	114.5	11.5	881	9	US-09-816-860A-2	Sequence 2, Appl
44	114	11.5	487	16	US-10-437-963-137896	Sequence 137896,
45	113.5	11.4	305	15	US-10-425-114-70697	Sequence 70697, A

ALIGNMENTS

RESULT 1

US-09-939-853A-94  
; Sequence 94, Application US/09939853A  
; Publication No. US20040039163A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess et al.  
; TITLE OF INVENTION: No. US20040039163A1e1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-099  
; CURRENT APPLICATION NUMBER: US/09/939,853A  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 60/228,191  
; PRIOR FILING DATE: 2000-08-25  
; PRIOR APPLICATION NUMBER: 60/267,300  
; PRIOR FILING DATE: 2001-02-08  
; PRIOR APPLICATION NUMBER: 60/269,961  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/277,337  
; PRIOR FILING DATE: 2001-03-20  
; NUMBER OF SEQ ID NOS: 159  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 94  
; LENGTH: 658  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-939-853A-94

Query Match 93.8%; Score 933.5; DB 11; Length 658;

Best Local Similarity 67.5%; Pred. No. 3.1e-64; Mismatches 1; Indels 87; Gaps 1;

Matches 185; Conservative 1; Mismatches 1; Indels 87; Gaps 1;

Qy 1 LPVKKPLRGRPGSLRLA----- 21

Db 27 LPVKKPLRGRPGSLRLAEEVSGGTGLRSALSVPPPPQAGSRAGSGTGTGSDPPM 86

Qy 22 -----VSASPPRPLKE 33

```
Db      87 ERGAGAGKLPDTGRCVPTEGSTVOLTIAPWNAADVHSHGDXDSQTCIRVSASDPDPRIKE 146
Qy      34 EEEAPLLPRTHLOAEPHGHGQWTVTEPAAMTPGNTTTPRTPEVTPPLRLLEQLKPLGLASTT 93
Db      147 EEEAPLLPRTHLOAEPHGHGQWTVTEPAAMTPGNTTTPRTPEVTPPLRLLEQLKPLGLANTT 206
Qy      94 LSTNPDTQASASDPDRPLRBEERARLLPRTHLOAELHGHGQWTVTEPAALTCGNATPPR 153
Db      207 LSTNPDTQASASDPDRPLRBEERARLLPRTHLOAELHGHGQWTVTEPAALTCGNATPPR 266
Qy      154 TQEVTPLELLEQLKPLPELVHATLSTPNPDNQVTIK 187
Db      267 TQEVTPLELLEQLKPLPELVHATLSTPNPDNQVTIK 300

RESULT 2
US-10-074-978A-403
; Sequence 403, Application US/10074978A
; Publication No. US20040010119A1
; GENERAL INFORMATION:
; APPLICANT: Leite, Mario
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Fernandes, Elma
; APPLICANT: Li, Li
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Liu, Xiaohong
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Patturajan, Meera
; APPLICANT: Blalock, Angela
; APPLICANT: Ballinger, Robert
; APPLICANT: Vernet, Corine
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Gusev, Vladimir
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter S
; APPLICANT: Ellerman, Karen
; APPLICANT: Heyes, Melvin P
; APPLICANT: Herrman, John
; APPLICANT: Pena, Carol E A
; APPLICANT: Shinkets, Richard A
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Moore, No. US20040010119A1lle
; APPLICANT: Shency, Suresh
; APPLICANT: Edinger, Shlomit
; APPLICANT: Gunther, Erik
; APPLICANT: Stone, Dave
; APPLICANT: Millet, Isabelle
; APPLICANT: Feyman, John
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-269
; CURRENT APPLICATION NUMBER: US/10/074,978A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: 60/268,221
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/335,109
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/312,284
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/268,496
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 60/276,703
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/330,293
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/322,127
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/280,899
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/310,797
```

```

; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/268,646
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 403
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-074-978A-403

Query Match      88.3%; Score 879; DB 15; Length 538;
Best Local Similarity 98.8%; Pred. No. 4.1e-60;
Matches 164; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      22 VSASDPDRPLKREBEAPLLPRTHLOAEPHGHGQWTVTEPAAMTPGNTTTPRTPEVTPPLRL 81
Db      15 VSASDPDRPLKREBEAPLLPRTHLOAEPHGHGQWTVTEPAAMTPGNTTTPRTPEVTPPLRL 74
Qy      82 ELQKLPLGLASTTLPSTNPDTQASASDPDRPLRBEERARLLPRTHLOAELHGHGQWTVTEP 141
Db      75 ELQKLPLGLANTTLPSTNPDTQASASDPDRPLRBEERARLLPRTHLOAELHGHGQWTVTEP 134
Qy      142 AALTTPGNATPPRTQEVTPLELLEQLKPLPELVHATLSTPNPDNQVTIK 187
Db      135 AALTTPGNATPPRTQEVTPLELLEQLKPLPELVHATLSTPNPDNQVTIK 180

RESULT 3
US-09-864-761-33474
; Sequence 33474, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/532,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
```

;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 33474  
;; LENGTH: 81  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AF11168.2  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.85  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.82  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.75  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.92  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.96  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.97  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4  
;; OTHER INFORMATION: EST\_HUMAN HIT: BE281519.1, EVALUATE 1.30e+00  
;; OTHER INFORMATION: EST\_HUMAN HIT: BE731965.1, EVALUATE 2.00e-31  
;; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUATE 5.30e+00  
US-09-864-761-33474

Query Match 43.2%; Score 430; DB 9; Length 81;  
Best Local Similarity 97.5%; Pred. No. 3.4e-26;  
Matches 79; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 22 VSASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMTGCNTTPTTPEVTPLRL 81  
Db 1 VSASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMTGCNTTPTTPEVTPLRL 60  
  
Qy 82 ELQKLPLGLASTLSTPNPDQ 102  
Db 61 ELQKLPLGLANTLSTPNPDQ 81

RESULT 4  
US-10-124-557-14  
; Sequence 14, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Bridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991

;; APPLICATION NUMBER: US 07/546,114  
;; FILING DATE: 29-JUN-1990  
;; APPLICATION NUMBER: US 07/457,196  
;; FILING DATE: 29-DEC-1989  
;; APPLICATION NUMBER: US 07/390,901  
;; FILING DATE: 08-AUG-1989  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Cserit, Luann  
;; REGISTRATION NUMBER: 31,822  
;; REFERENCE/DOCKET NUMBER: GI 5190  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617)876-1170  
;; TELEFAX: (617)876-5851  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 941 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-10-124-557-14  
  
Query Match 14.5%; Score 144; DB 13; Length 941;  
Best Local Similarity 28.4%; Pred. No. 0.0083;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;  
  
Qy 23 SASDPRLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71  
Db 356 TTTKEPATTTPKEPATTTPKPAATTPKEPATTTPKPAATTPKPAATTPKPAATTPK 415  
  
Qy 72 RTPEVTPLRLQKLPLGLASTLSTPNPDQASAP-----DPRPLREERARLLPRTHLQ 127  
Db 416 ETAPTTPKLTPTTPEKLAATTPKEPATTTPPEELAPTTPEPTPTTPEPAPTTPKAAAP 475  
  
Qy 128 AELHQHCWTVTEPAALT---PGNATPRTQEVTPLLLELQKLPLVHATLSTPNP 180  
Db 476 NTPKEPATTTPKEPATTTPKPAATTPKPAATTPKPAATTPKPAATTPKPAATTPK 531

RESULT 5  
US-10-124-557-84  
; Sequence 84, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Bridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196

```

; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1022 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84
;
; Query Match 14.5%; Score 144; DB 13; Length 1022;
; Best Local Similarity 28.4%; Pred. No. 0.0091;
; Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;
;
; QY 23 SASDPRLKKEEAPLLPRHLQAEPHQHGCWTVTEPAAMT-----PGNTTPP 71
; DB 437 TTTKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 496
;
; QY 72 RTPEVTLRLLEQLKPLGLASTTSTNPNDTQASASP-----DPRPLREEEARLLPRHLQ 127
; DB 497 ETAPTTKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEELAPTTPEELAPTT 556
;
; QY 128 AELHQHGCWTVTEPAALT---PGNATPRTQEVTLLEQLKPLVHATLSTNP 180
; DB 557 NTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTT 612
;
; RESULT 6
; US-10-124-557-74
; Sequence 74, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
;
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1038 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74
;
; Query Match 14.5%; Score 144; DB 13; Length 1038;
; Best Local Similarity 28.4%; Pred. No. 0.0093;
; Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;
;
; QY 23 SASDPRLKKEEAPLLPRHLQAEPHQHGCWTVTEPAAMT-----PGNTTPP 71
; DB 421 TTTKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 480
;
; QY 72 RTPEVTLRLLEQLKPLGLASTTSTNPNDTQASASP-----DPRPLREEEARLLPRHLQ 127
; DB 481 ETAPTTKKLTPTTPEKLAPTTPEKAPTTPEELAPTTPEELAPTTPEELAPTT 540
;
; QY 128 AELHQHGCWTVTEPAALT---PGNATPRTQEVTLLEQLKPLVHATLSTNP 180
; DB 541 NTPKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 596
;
; RESULT 7
; US-10-124-557-58
; Sequence 58, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822

```

```
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match          14.5%; Score 144; DB 13; Length 1049;
Best Local Similarity 28.4%; Pred. No. 0.0094;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASPPRLKKEEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 464 TTKKEPATTTPKEPATTTPKKPAPTTPKEPATTTPKEPATTTPKKPAPTTPK 523
Qy 72 RTPEVTLRLLEQLKPLGLASTTLPNDPDTQASAP-----DPRPLREEEEARLLPRTHLQ 127
Db 524 ETAPTPKLTPTTPEKLAPTTPKEPATTTPKEPATTTPPEELAPTTPPEPTPTTPEEPATTPKAAP 583
Qy 128 AELHQHCWTVTEPAALT---PGNATPRTQEVTPLLLEQLKPLVHATLSTNP 180
Db 584 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEPATTTPKKAP 639

RESULT 8
US-10-124-557-104
; Sequence 104, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceerr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

Query Match          14.5%; Score 144; DB 13; Length 1140;
Best Local Similarity 28.4%; Pred. No. 0.01;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;

Qy 23 SASPPRLKKEEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 555 TTKKEPATTTPKEPATTTPKKPAPTTPKEPATTTPKEPATTTPKKPAPTTPK 614
Qy 72 RTPEVTLRLLEQLKPLGLASTTLPNDPDTQASAP-----DPRPLREEEEARLLPRTHLQ 127
Db 615 ETAPTPKLTPTTPEKLAPTTPKEPATTTPKEPATTTPPEELAPTTPPEPTPTTPEEPATTPKAAP 674
Qy 128 AELHQHCWTVTEPAALT---PGNATPRTQEVTPLLLEQLKPLVHATLSTNP 180
Db 675 NTPKEPATTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEPATTTPKKAP 730

RESULT 9
US-10-124-557-44
; Sequence 44, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceerr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
```

```

;
; LENGTH: 1270 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-124-557-44
Query Match 14.5%; Score 144; DB 13; Length 1270;
Best Local Similarity 28.4%; Pred. No. 0.012;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;
US-10-124-557-44
QY 23 SASPPRLKEEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 421 TTTKPEAPTTPEKPAATTPKPAATTPKEPAPTTPEKPAATTPKPAATTPKPAATTPK 480
QY 72 RTEVTPRLLEQLKPLGLASTTSLTPNPDQASASP-----DPRPLREEEEARLLPRTHLQ 127
Db 481 ETAPTTPKLTTTPEKLAPTTPEKPAATTPBELAPTTPEEPTPTTPEAPTTPKAAAP 540
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTQEVTPVLLLELQKLDELVHATLSTNP 180
Db 541 NTPKEPAPTTPEKPAATTPKPAATTPKETAATTPKGTAPTTLKEPAPTTPKKAP 596
RESULT 10
US-10-124-557-42
; Sequence 42, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Bridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-142
;
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42
Query Match 14.5%; Score 144; DB 13; Length 1311;
Best Local Similarity 28.4%; Pred. No. 0.012;
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;
US-10-124-557-42
QY 23 SASPPRLKEEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP 71
Db 462 TTTKPEAPTTPEKPAATTPKPAATTPKEPAPTTTKKPAATTPKPAATTPKPAATTPK 521
QY 72 RTEVTPRLLEQLKPLGLASTTSLTPNPDQASASP-----DPRPLREEEEARLLPRTHLQ 127
Db 522 ETAPTTPKLTTTPEKLAPTTPEKPAATTPBELAPTTPEEPTPTTPEAPTTPKAAAP 581
QY 128 AELHQHCWTVTEPAALT---PGNATPPRTQEVTPVLLLELQKLDELVHATLSTNP 180
Db 582 NTPKEPAPTTPEKPAATTPKPAATTPKETAATTPKGTAPTTLKEPAPTTPKKAP 637
RESULT 11
US-10-124-557-142
; Sequence 142, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Bridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

```



Query Match	14.5%;	Score 144;	DB 13;	Length 1313;
Best Local Similarity	28.4%;	Pred.No. 0.012;		
Matches	50;	Conservative 11;	Mismatches 97;	Indels 18; Gaps 3;

  

Qy	23	SASPDRLKEEAEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PQNTTPP	71
	:	: : : : : :	:
	:	: : : : : :	:
Db	464	TTTKPEAPPTPKGPAPTTPKKPAPPTTPKPEAPPTTKKPAPATPKBPAPTPFK	523
	:	: : : : : :	:
Qy	72	RTEVPTPLRIELQKLPGLASTLTSTNPNPDQOASASP----DPRPLREEEARLLPRTHLQ	127
	:	: : : : : :	:
Db	524	ETAPPTPKLITPTPEKLAFTTEPKPAPTPEELAPTTPEEPTTTTEEPAPTPPKAAP	583
	:	: : : : : :	:
Qy	128	AELHQHCWTVTEBPAALT---PGNATPPRQTQEVPTPLLELKQLPELVHATLSTNP	180
	:	: : : : : :	:
Db	584	NTPKEAPPTPKGPAPPTTPKPEAPPTTPKETAPPTPKGTAPTTLKEPAPTPPKCAP	639
	:	: : : : : :	:

RESULT 12  
US-10-124-557-50  
; Sequence 50, Application US/10124557  
; Publication No. US20020137894A1

GENERAL INFORMATION:  
APPLICANT: Turner, Katherine  
Clark, Stephen C.  
Jacobs, Kenneth  
Hewick, Rodney M.  
Geaner, Thomas G.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

; CURRENT AFFILIATION DATA:  
 ; APPLICATION NUMBER: US/10/124,557  
 ; FILING DATE: 16-Apr-2002  
 ;

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:

NAME: CSEILL, LUDMIL  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 876-1170  
TELEFAX: (617) 876-5851

```

; INFORMATION FOR SEQ ID NO: 50:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1314 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;
; MOLECULE TYPE: protein

```

US-10-124-557-50

Query Match 14.5%; Score 144; DB 13; Length 1314;  
Best Local Similarity 28.4%; Pred. No. 0.012;

Matches	50; Conservative	11; Mismatches	97; Indels	18; Gaps	3;
QY	23 SASDPRLPKEEAEALLPRTHLQAEPHQHCWTVTEPAAMT-----PGNTTTP	71			
Db	465 TTTKEPATTTPKEPATTTPKCAPATTTPKEPATTTPKCAPATTTPKCAPATTTPK	524			
QY	72 RTPEVTPRLLEQLKPLGLASTTTLSTNPDTQASASP-----DPRPLREEEARLLPRTHLQ	127			
Db	525 ETATTPPKKLTTPTEKLAATPEKAPATTPEBLATTPPEPTPTTPEEPATTPPKAAAP	584			
QY	128 AEHLQHGCWTVTEPAALT---PGNATPPRTQEVTPPLLLLEQLKPELVHATLSTNP	180			
Db	585 NTPKEPATTTPKEPATTTPKEPATTTPKCAPATTTPKGTATTTLKEPATTTPKCAP	640			

RESULT 13  
US-10-124-557-46  
; Sequence 46, Application US/10124557  
; Publication No. US20020137894A1

GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 NUMBER OF SEQUENCES: 143  
 CORRESPONDENCE ADDRESS: -  
 ADDRESSEE: Genetics Institute, Inc.  
 STREET: 87 CambridgePark Drive  
 CITY: Cambridge  
 STATE: Massachusetts  
 COUNTRY: U.S.A.  
 ZIP: 02140

ZIT: UZ14U  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0,  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002

/ BASIC DATA: /  
 / PRIORITY: /  
 / PRIOR APPLICATION DATA: /  
 / APPLICATION NUMBER: US 07/643,502 /  
 / FILING DATE: 18-JAN-1991 /  
 / APPLICATION NUMBER: US 07/546,114 /  
 / FILING DATE: 29-JUN-1990 /  
 / APPLICATION NUMBER: US 07/457,196 /  
 / FILING DATE: 29-DEC-1989 /  
 / APPLICATION NUMBER: US 07/390,901 /  
 / FILING DATE: 08-AUG-1989 /  
 /

NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170

TELEFAX: (017) 878-3831  
 INFORMATION FOR SEQ ID NO: 46:  
 SEQUENCE CHARACTERISTICS:

LENGTH: 1320 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-10-124-557-46

Query Match 14.5%; Score 144; DB 13; Length 1320;  
Best Local Similarity 28.4%; Pred. No. 0.012;  
Matches 50; Conservative 11; Mismatches 97; Indels 18

QY 23 SASPDPRPLKEEEEEAPLPRTHLQAEPHQHGCWTVTEPAAMT-----PGNTTPP 71



Search completed: February 19, 2005, 00:11:41  
Job time : 134 secs

**This Page Blank (uspto)**





C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Jul-2004  
C/Accession: S3615; S75613; S27721  
R/Malakhov, M.P.; Wada, H.; Los, D.A.; Sakamoto, T.; Murata, N.  
Plant Mol. Biol. 21, 913-918, 1993  
A>Title: Structure of a cyanobacterial gene encoding the 50S ribosomal protein L9.  
A/Reference number: S33614, UID:93222488; PMID:8467083  
A/Accession: S33615  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-291 <NA>  
A/Cross-references: UNIPROT:P42350; GB:D10716; NID:g217098; PIDN:BAA38818.1; PID:d100203  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, April 1992  
R/Kaneko, T.; Sato, S.; Kocani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shampo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.

A/Reference number: S74322; UID:97061201; PMID:8905231  
A/Accession: S75613  
A>Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-291 <KAN>  
A/Cross-references: EMBL:D90912; GB:AB001339; NID:g1653228; PIDN:BAA18174.1; PID:d101890  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match            11.7%     Score 116; DB 2; Length 291;  
Best Local Similarity      26.8%; Pred. No. 0.2;  
Matches       49; Conservative    21; Mismatches    61; Indels    52; Gaps    9;

QY          13 PGSLRLAEVSAPDRPLKEEAPLLPRTHLOAPHQHGCWTVTEPAAMTPGNTTPPR 72  
Db           |::| ::|| ||:|| :|::| :|::| :|::| :|::| :|:  
83 PTAIVDEMAKPSPPSPKKESKP-----KONHK---VVTFPAVNPPTVPVPAH 130  
  
QY          73 TPETVLRLLEQLKL----GLASTTLSTPNPDQAASDPDLREEEEARLLPRTLQA 128  
Db           |:|:| :|::| :|::| :|::| :|::| :|::| :|::| :|::| :|:  
131 PP--TPV---VEKSPEVEAIAIEPIITRAP-----ISFPSPDVLSEEP----- 172  
  
QY          129 ELRHQCWTVTPEPALTPGNATPRTQEVTPLLELQ-----KLDELVHA TLSTPNPDN 182  
Db           |::|:|:| :|:|:|:| :|:|:|:| :|:|:|:| :|:|:|:| :|:|:|:| :|:  
173 -----TPPPAVNV-NSTNQPEESAVIDSELQDFATPELPFLAVEAKPDSPEPDM 220  
  
QY          183 QVT 185  
Db           |:  
221 AVS 223

RESULT 7  
T30826  
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse  
N/Alternate names: alpha-NAC protein  
C/Species: Mus musculus (house mouse)  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T30826  
R/Yotov, W.V.; St-Arnaud, R.  
Genes Dev. 10, 1763-1772, 1996  
A>Title: Differential splicing-in of a proline-rich exon converts alphanac into a muscle  
A/Reference number: Z20889; UID:96312450; PMID:8698236  
A/Accession: T30826  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-2187 xYT>  
A/Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB187  
C/Genetics:  
A/Gene: Naca  
A/Map position: 10  
A/Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3  
A>Note: differential splicing converts alphanac into a tissue-specific DNA-binding activ  
C/key words: alternative splicing; DNA binding; transcription factor

Query Match            11.6%     Score 115.5; DB 2; Length 2187;  
Best Local Similarity      20.5%; Pred. No. 1.9;  
Matches       56; Conservative    32; Mismatches    96; Indels    89; Gaps    8;

A:Molecule type: DNA  
A:Residues: 1-1186 <WIL>  
A:Cross-references: UNIPROT:Q17786; EMBL:Z49908; PIDN:CAA90094.1; GSPDB:GN00020; CESP:CO  
A:Experimental source: clone C07E3  
C:Genetics:  
A:Gene: CESP:C07E3.3  
A:Map position: 2  
A:Introns: 36/3; 79/3; 1128/3; 1160/3

Query Match 11.5%; Score 114; DB 2; Length 1186;  
Best Local Similarity 27.3%; Pred. No. 1.2;  
Matches 54; Conservative 19; Mismatches 71; Indels 54; Gaps 11;

QY 3 VKKRLRGPRGSLRLAEVSAASDPRLKKEERAPLLPRTHLQAEPHQHCWTVTEPA- 61  
DB 694 VASPRRYTPP-----SSPKXHLTHLREVEFGTPLNHYIGRPRSTPIYS-TPPED 743  
QY 62 --AMTPTGNTTPRTPEV-----TPLRLLEQLKLP-----GLASTTLSTPNPDTOAS 104  
DB 744 RNVVSPSRVTPP--FSVTMAAKLHLTLQPEVPSTPLQHYISRSASSVYNTPSVD---- 797  
QY 105 ASDPRPLRBEERALLRTHLQAEHLQHGCWT-----VTEPAALTPGNATPPRTQVETPL 161  
DB 798 -STSTRSLVSPSSVRMPKHLH-TRLQDVPWTPFKVSRPRS-TPITYSTPP----- 845  
QY 162 LLEQLKLPVHATLSTPN 179  
DB 846 -----LEKPISTLVTPS 859

RESULT 10  
A43932  
mucin 2 precursor, intestinal - human (fragments)  
N:Alternate names: mucin SMUC-41  
C:Species: Homo sapiens (man)  
C:Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004  
A:Accession: A49963; A45106; A43932; B33532; A61257; PQ0328; PQ0329  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the  
A:Reference number: A49963; MUID:94132002; PMID:8300571  
A:Accession: A49963  
A:Molecule type: mRNA  
A:Residues: 1-639 <GUI>  
A:Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:L21998  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr  
A:Reference number: A45106; MUID:93016075; PMID:1400449  
A:Accession: A45106  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 626-1895 <GU2>  
A:Cross-references: GB:M94131; NID:gl86395; PIDN:AAA59163.1; PID:gl86396  
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A:Accession: B45106  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 2037-3020 <GU3>  
A:Cross-references: GB:M94132; NID:gl86397; PIDN:AAA59164.1; PID:gl86398  
A:Experimental source: colon  
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M  
J. Clin. Invest. 88, 1005-1013, 1991  
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp  
A:Reference number: A49932; MUID:9135817; PMID:1885763  
A:Accession: A49932  
A:Molecule type: DNA  
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; NID:gl88863; PIDN:AAA59875.1; PID:gl88864  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIP:55749, NCBIP:55750)  
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.

J. Biol. Chem. 264, 6480-6487, 1989  
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden  
A:Reference number: A33532; MUID:89197956; PMID:2703501  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GU4>  
A:Cross-references: GB:M22405; NID:gl88873; PIDN:AAA36334.1; PID:gl88874  
A:Experimental source: intestine  
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.  
J. Clin. Invest. 87, 77-82, 1991  
A:Title: Human bronchus and intestine express the same mucin gene.  
A:Reference number: A61257; MUID:91086481; PMID:1985113  
A:Accession: A61257  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>  
A:Experimental source: bronchus  
R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner,  
Biochem. Biophys. Res. Commun. 183, 821-828, 1992  
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-t  
A:Reference number: PQ0328; MUID:92198477; PMID:1550588  
A:Accession: PQ0328  
A:Molecule type: mRNA  
A:Residues: 2328-2468 <XUG>  
A:Cross-references: GB:M86523  
A:Experimental source: small intestine  
A:Accession: PQ0329  
A:Molecule type: protein  
A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>  
C:Genetics:  
A:Gene: GDB:MUC2  
A:Cross-references: GDB:I20203; OMIM:158370  
A:Map position: 11p15.5-11p15.5  
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von  
C:Keywords: glycoprotein; intestine; tandem repeat  
F:2766-2834/Domain: von Willebrand factor type C repeat homology <VMC>

Query Match 11.4%; Score 113; DB 2; Length 3020;  
Best Local Similarity 26.2%; Pred. No. 4;  
Matches 48; Conservative 14; Mismatches 99; Indels 22; Gaps 7;

QY 13 PGSRLAEVSAASDP---RPLKEEERAPLLPRTHLQAEPHQHCWTVTEPAAMTPG--N 67  
DB 1526 PTTTTTTPPTTTPSPPTTTPITPTTSTTLPLPTTSPSP---PTTTTTPPTTTPSPPT 1582  
QY 68 TTPRTTEVPLRLLEQLKLPCLASTTLSTPNPDQASDPDRPLREERARLLPRTHLQ 127  
DB 1583 TTTPSPPTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPS-PPTTTPPTTTPPTTTP 1641  
QY 128 AELHQHCWTVTEPAALTPGNAT-----PPRTQVETP-----LLEQLKLPVHATLST 177  
DB 1642 S---PPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1698  
QY 178 PNP 180  
DB 1699 PSP 1701

RESULT 11  
Q068  
BPLF1 protein - human herpesvirus 4 (strain B95-8)  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C:Date: 25-Feb-1995 #sequence revision 25-Feb-1985 #text change 09-Jul-2004  
A:Accession: G93065; A03747; S32993  
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-45, 1983  
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus  
A:Reference number: A93065; MUID:85035713; PMID:6092825  
A:Accession: G93065  
A:Molecule type: DNA  
A:Residues: 1-3149 <BAN>  
A:Cross-references: UNIPROT:P03186; EMBL:V01555; NID:959074; PIDN:CAA24839.1; PID:gl33348  
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.;



Nature 310, 207-211, 1984  
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667; PMID:6087149  
A:Contents: annotation; protein coding region  
C:Superfamily: human herpesvirus 4 BPLF1 protein

Query Match 11.3%; Score 112; DB 1; Length 3149;  
Best Local Similarity 25.6%; Pred. NO. 5;  
Matches 57; Conservative 21; Mismatches 79; Indels 66; Gaps 12;

QY 11 PRGSLTRLAEVSASDPDR-----PLKE---EEAPLLPRTHLQAE 48  
DB 395 PAPSTPRASSGAADQTPRKKGKGDSPHKKPGSGRLPLSSTTDEDDQLPRTHVDPH 454  
QY 49 -----PHQHGCTVTPEAAMTGNTPPT--PEVTPL-RLELQ---KLP 87  
DB 455 RPPSAARLPPVPIPIHQ-----SPPASPTP-HPAPVSTIAPSVTPSPRLPLQIPILP 507  
QY 88 GLASTT-----LSTNPDTQASASP-----DRPLREEEEARLLPRTHLQAEIHQHGCTV 138  
DB 508 QAAPSNPKIPLTTPSPSPTAAAPTTLTSPPTQQQPPQSAAPAPSPLLPQQQPTPSAA 567  
QY 139 TEPAALTPGNATPPTQEV-TPLLELQKLPELVHATLSTNP 180  
DB 568 PAPGELLPQQQPPSAARAPSLPPPPQQPLPS-----ATPAP 604

RESULT 12  
E70546  
hypoethical protein RV0538 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: E70546  
R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: E70546  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-548 <COL>  
A:Cross-references: UNIPROT:O06404; GB:Z95558; NID:G3261781; PIDN:CAB08989.  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV0538

Query Match 11.2%; Score 111.5; DB 2; Length 548;  
Best Local Similarity 26.6%; Pred. NO. 0.82;  
Matches 46; Conservative 10; Mismatches 84; Indels 33; Gaps 5;

QY 11 PRGSLTRLAEVSASDPDRPLKEEEAPLLPRTHLQABP-----HQHGCTVT 58  
DB 360 PRPGTSGVGCTPASAP-----EAPAPGVVPAPVPIPIIIPFPQWQGMETIP 412  
QY 59 EPAAMTP--GNTTPPTPEVTPLRLQKLGASTTLSTNPDTQASAPDRPLREE 115  
DB 413 TAPPTTPVTTTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPV 472  
QY 116 EARLLPRTHLQAEIHQHGCTVTTEPAALTPGNATPPTQEV-TPLLELQKL 168  
DB 473 APTTVAPTTPV-----TTVAPTATTTTVAPOPTQO--PFOQPTQMP 514

RESULT 13  
T21700  
hypoethical protein F33B2.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T21700  
R:Lennard, N.

submitted to the EMBL Data Library, January 1997  
A:Reference number: Z19461  
A:Accession: T21700  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-846 <WIL>  
A:Cross-references: UNIPROT:O01699; EMBL:Z84574; PIDN:CAB06541.1; GSPDB:GN00019; CESP:F3  
A:Experimental source: clone F33E2  
C:Genetics:  
A:Gene: CESP:F33E2.6  
A:Map position: 1  
A:Introns: 99/3; 213/3; 332/3; 525/3

Query Match 11.2%; Score 111; DB 2; Length 846;  
Best Local Similarity 24.9%; Pred. NO. 1.4;  
Matches 47; Conservative 20; Mismatches 92; Indels 30; Gaps 7;

QY 6 PRLGPRPGSLTRLAEVSASDPDRPLKEEEAPLL--PRTHLQAEPHQHGCTVTTEPAAM 63  
DB 571 PRTEPPR-----TEPPKTEAPRTVRPKTEAPMTVPPTTEPPMTEAPRTVEVPMTPEPKT 623  
QY 64 TPGNTTPTRT-----PEVTPLRLQKLGASTTLSTNPDTQASAPDRPLREE 116  
DB 624 EPPRTAPRTVEVSMPLPPEVTPNTEAPR-----TEVMTVP-PRTEPPKTEAPRTVPPKT 678  
QY 117 EARLLPRTHLQAEIHQHGCTVTTEPAALTPGNATPPTQEV-TPLLELQKLPELVHATLS 176  
DB 679 EA---PWTEVMTGPTSEVMTPEPKTEQPTAPRTE-----VSMPLPPEVTPPKTE 729  
QY 177 TPNPDNQVT 185  
DB 730 APRTVEPMT 738

RESULT 14  
A45690  
transactivator EBNA-2 - baboon lymphotropic virus  
C:Species: baboon lymphotropic virus, herpesvirus papio HVP  
C:Date: 21-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: A45690  
R:Ling, P.D.; Ryon, J.J.; Hayward, S.D.  
J. Virol. 67, 2990-3003, 1993  
A:Title: EBNA-2 of herpesvirus papio diverges significantly from the type A and type B E  
ic motif.  
A:Reference number: A45690; MUID:93267743; PMID:8388484  
A:Accession: A45690  
A:Molecule type: DNA  
A:Residues: 1-530 <LIN>  
A:Cross-references: UNIPROT:Q07701; GB:L11366; NID:G6690013; PIDN:AAA79034.1; PID:G30631  
A:Note: sequence extracted from NCBI backbone (NCBIN:132361, NCBIPI:132362)  
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 11.0%; Score 109.5; DB 2; Length 530;  
Best Local Similarity 22.4%; Pred. NO. 1.1;  
Matches 56; Conservative 25; Mismatches 92; Indels 77; Gaps 11;

QY 3 VKPR-----LRGPRGSLTRLAEVSASDPDRPLKEEEAPLLPRTHLQASPH 50  
DB 151 VPRPRMTILPLRLIRPRPTYNPLSPNTSPSPPIF---QHTPLPRPTASPH 207  
QY 51 OHGC-----WTVTEPAAMTGN--TTPPTPEVTPLRLQKLGAST----TLSTNP 99  
DB 208 SHPQSPKQTPPKPTLPLGPPVSPPPPTPSIQTPHTPPKPTIFSSGPHITL-TVDP 266  
QY 100 -----DTQASAPDRPLREE-----EARLLPRTHLQAEIHQHGCTVTE 140  
DB 267 LSQPASTQGSNVSPQPFQHKQILITTLATSOPTTKQILPKTRSS-----ASMD 318  
QY 141 PAALTGPNATPPTQEV-----TPLLLELQKL--PELVHATLST 177  
DB 319 PLPLPPLSTPPPPAPSTPSGIVDRDRTSPRLGPNVWPPVLPFHKLAGDILLTPSPDP 378  
QY 178 PNPDNQVTIK 187

```

| : : |
Db      379 PTPEETVRK 398

RESULT 15
C87437
penicillin-binding protein, 1A family [imported] - Caulobacter crescentus
CtSpecies: Caulobacter crescentus
CtDate: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
CtAccession: C87437
CtRnierman, W.C.; Feidblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolor
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87437
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-794 <STO>
A;A;Cross-references: UNIPROT:Q9A848; GB:AE005673; NID:gl3422895; PIDN:AAK23495.1; GSPDB:G
C;Genetics:
A;Gene: CC1516

```

Query Match	11.0%;	Score	109.5;	DB	2;	Length	794;
Best Local Similarity	27.1%;	Pred. No.	1.7;				
Matches	49;	Conservative	18;	Mismatches	79;	Indels	35;
						Gaps	10;

  

QY	2	PVKKPLRGRPGSLRLA	SVSPDPRLK	EEEEAPLLPR	THLOAEPHQHCW	VTTEPA	61
Db	15	PAPKPR-RPKPRAQT	WMTZ-AGLDPA	EDAAQKPRPK	SRRAQPE-----	AVAAPE	67
QY	62	AMTPGNNTTPRTPE	VTPLRLLEQKL	PGLASTLTST	PNPDQO-ASASPD	PRPLREEPEARL	120
Db	68	AQPTTSTAATDADA	DL-----FB	PLP-----SP	PDAPSVVEADPA	ETLEADAA--	114
QY	121	LPRTHLOAEHLQH	QHCWVTTEPA	ALT	PGNATPPRTQ	EVTPLLLEQLQKL	PELVHATLSTPNP
Db	115	-----ADLE-	-SPT	EP-EVEPAQDIP	PEAPPEPTTA	EADAPE-AKAETSV	PVTL
QY	181	D	181				
Db	161	D	161				

Search completed: February 18, 2005, 23:59:49  
Job time : 42 secs



DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Hypothetical protein DKFZp686E0215 (Fragment).  
 GN Name=DKFZp686E0215;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cervix;  
 RG The German cDNA Consortium;  
 RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,  
 RA Fobo G., Han M., Wiemann S.;  
 RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; CR749863; CAH18707.1; -;  
 DR InterPro; IPR005533; AMOP.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF03782; AMOP; 1.  
 DR Pfam; PF00090; TSP 1; 1.  
 DR SMART; SM00723; AMOP; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 KW Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 578 AA; 64611 MW; 4BA9FA46801BD68B CRC64;  
 Query Match 99.7%; Score 992; DB 2; Length 578;  
 Best Local Similarity 99.5%; Pred. No. 2.3e-58;  
 Matches 186; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEAPLLPRTHLQAEPHQHCWTVTEP 60  
 DB 34 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEAPLLPRTHLQAEPHQHCWTVTEP 93  
 QY 61 AAMTPGNTTTPRTPEVTPRLRLQKPLGLASTTLLSTNPDTQASASDPRLREEEEARL 120  
 DB 94 AAMTPGNTTTPRTPEVTPRLRLQKPLGLANTTLLSTNPDTQASASDPRLREEEEARL 153  
 QY 121 LPRTHLQAEHLHQHCWTVTEPAALTPGNATPPRTQEVTPLLLELQKLPVHATLSTNP 180  
 DB 154 LPRTHLQAEHLHQHCWTVTEPAALTPGNATPPRTQEVTPLLLELQKLPVHATLSTNP 213  
 QY 181 DNQVTIK 187  
 DB 214 DNQVTIK 220  
 RESULT 3  
 Q86TQ7 PRELIMINARY; PRT; 328 AA.  
 AC Q86TQ7  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Full-length cDNA 5-PRIME end of clone CS0DI033VJ09 of Placenta of Homo sapiens (human) (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Li W.B., Gruber C., Jessee J., Polayes D.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Genoscope;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BX248770; CAD66577.1; -;

FT NON TER 1  
 SQ SEQUENCE 328 AA; 36307 MW; 56C700854F62E89B CRC64;  
 Query Match 99.2%; Score 987; DB 2; Length 328;  
 Best Local Similarity 98.9%; Pred. No. 2.8e-58;  
 Matches 185; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEAPLLPRTHLQAEPHQHCWTVTEP 60  
 DB 31 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEAPLLPRTHLQAEPHQHCWTVTEP 90  
 QY 61 AAMTPGNTTTPRTPEVTPRLRLQKPLGLASTTLLSTNPDTQASASDPRLREEEEARL 120  
 DB 91 AAMTPGNTTTPRTPEVTPRLRLQKPLGLANTTLLSTNPDTQASASDPRLREEEEARL 150  
 QY 121 LPRTHLQAEHLHQHCWTVTEPAALTPGNATPPRTQEVTPLLLELQKLPVHATLSTNP 180  
 DB 151 LPRTHLQAEHLHQHCWTVTEPAALTPGNATPPRTQEVTPLLLELQKLPVHATLSTNP 210  
 QY 181 DNQVTIK 187  
 DB 211 DNQVTIK 217  
 RESULT 4  
 Q6H9L7 PRELIMINARY; PRT; 571 AA.  
 ID Q6H9L7  
 AC Q6H9L7  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Thrombospondin and AMOP containing isthmin-like 1 protein, (Tail1) precursor.  
 GN Name=TAIL1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Heart;  
 RA Rossi V., Belfagna G., Rampazzo A., Baucé B., Danieli G.A.;  
 RT "TAIL1: an isthmin-like gene, containing type 1 thrombospondin-repeat and AMOP domain, mapped to ARVD1 critical region.";  
 RL Gene 335.101-108(2004).  
 DR EMBL; AJ583024; CAB47313.1; -;  
 DR InterPro; IPR005533; AMOP.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF03782; AMOP; 1.  
 DR Pfam; PF00090; TSP 1; 1.  
 DR SMART; SM00723; AMOP; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 KW Signal.  
 FT CHAIN 1 26 Potential.  
 FT SIGNAL 27 571 Tail1 protein.  
 SQ SEQUENCE 571 AA; 63906 MW; D83329D10F300C7C CRC64;  
 Query Match 99.2%; Score 987; DB 2; Length 571;  
 Best Local Similarity 98.9%; Pred. No. 5e-58;  
 Matches 185; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEAPLLPRTHLQAEPHQHCWTVTEP 60  
 DB 27 LPVKKPRLRGPRGSLRLAEVSASDPRLPKKEEAPLLPRTHLQAEPHQHCWTVTEP 86  
 QY 61 AAMTPGNTTTPRTPEVTPRLRLQKPLGLASTTLLSTNPDTQASASDPRLREEEEARL 120  
 DB 87 AAMTPGNTTTPRTPEVTPRLRLQKPLGLANTTLLSTNPDTQASASDPRLREEEEARL 146  
 QY 121 LPRTHLQAEHLHQHCWTVTEPAALTPGNATPPRTQEVTPLLLELQKLPVHATLSTNP 180

Db 147 LPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRTQEVTPLELLELQKLPVLVHATLSTNP 206  
 Qy 181 DNQVTIK 187  
 |||||  
 Db 207 DNQVTIK 213  
 |||||  
 RESULT 5  
 Q86TW3 PRELIMINARY; PRT; 294 AA.  
 AC Q86TW3;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Full-length cDNA clone CSOD1014YN05 of Placenta of Homo sapiens  
 (human) (Fragment).  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Li W.B., Gruber C., Jessee J., Polayes D.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Genoscope;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BX248277; CAD62605.1; -;  
 FT NON TER 1  
 SQ SEQUENCE 294 AA; 31563 MW; 5127610A07112FC2 CRC64;  
 Query Match 97.4%; Score 969; DB 2; Length 294;  
 Best Local Similarity 98.9%; Pred. No. 3.9e-57;  
 Matches 181; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 LPVKKPRLGRPRGSLTRLAEVSPDPRLKEEEAPLLPRTHLQAEHPHOGHCWTTVEP 60  
 |||||  
 Db 29 LPVKKPRLGRPRGSLTRLAEVSPDPRLKEEEAPLLPRTHLQAEHPHOGHCWTTVEP 88  
 |||||  
 Qy 61 AAMTPGNATPRTPEVTPLELLELQKLPGLASTTLSTNPDTQASASDPDPRLKEEEEARL 120  
 |||||  
 Db 89 AAMTPGNATPRTPEVTPLELLELQKLPGLANTTLSTNPDTQASASDPDPRLKEEEEARL 148  
 |||||  
 Qy 121 LPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRTQEVTPLELLELQKLPVLVHATLSTNP 180  
 |||||  
 Db 149 LPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRTQEVTPLELLELQKLPVLVHATLSTNP 208  
 |||||  
 Qy 181 DNQ 183  
 |||||  
 Db 209 DNQ 211  
 |||||  
 RESULT 6  
 Q95432 PRELIMINARY; PRT; 658 AA.  
 AC Q95432;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Multimegabase Sequencing Group;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA UW Multimegabase Sequencing Group;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF111168; AAD09622.1; -;  
 DR Genew; HGNC:23176; THSD3.  
 DR InterPro; IPR005533; AMOP.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF03782; AMOP; 1.  
 DR Pfam; PF00090; TSP; 1; 1.  
 DR SMART; SMO0723; AMOP; 1.  
 DR SMART; SMO0209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 658 AA; 72519 MW; 7B1DF224645B300F CRC64;  
 Query Match 93.8%; Score 933.5; DB 2; Length 658;  
 Best Local Similarity 67.5%; Pred. No. 2.1e-54;  
 Matches 185; Conservative 1; Mismatches 1; Indels 87; Gaps 1;  
 Qy 1 LPVKKPRLGRPRGSLTRLAE-----VSASDPDPRLKE 33  
 |||||  
 Db 27 LPVKKPRLGRPRGSLTRLAEVSGGTGLRSALSVPPPQFAGSSRAGSGTGTHGSDPPM 86  
 |||||  
 Qy 22 -----VSASDPDPRLKE 33  
 |||||  
 Db 87 ERGAGAGKLPDTCRCVPTGEGTVQLIAPNAAADVHSHGDKDSQTCIRVSASDPDPRLKE 146  
 |||||  
 Qy 34 EERAPLLPRTHLQAEHPHOGHCWTTVEPAAMTPGNATPRTPEVTPLELLELQKLPGLASTT 93  
 |||||  
 Db 147 EERAPLLPRTHLQAEHPHOGHCWTTVEPAAMTPGNATPRTPEVTPLELLELQKLPGLANTT 206  
 |||||  
 Qy 94 LSTNPDTQASASDPDPRLKEEEEARLLPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRT 153  
 |||||  
 Db 207 LSTNPDTQASASDPDPRLKEEEEARLLPRTHLQAEHLHOGHCWTTVEPAALTPGNATPRT 266  
 |||||  
 Qy 154 TOEVTPLLELQKLPVLVHATLSTNPDPNQVTIK 187  
 |||||  
 Db 267 TOEVTPLLELQKLPVLVHATLSTNPDPNQVTIK 300  
 |||||  
 RESULT 7  
 Q8N501 PRELIMINARY; PRT; 518 AA.  
 ID Q8N501  
 AC Q8N501;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE THSD3 protein (Fragment).  
 GN Name=THSD3;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
[2]

RC SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC033140; AAH33140.1; -;  
DR InterPro; IPR005533; AMOP.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF03782; AMOP; 1.  
DR Pfam; PF00090; TSP 1; 1.  
DR SMART; SM00723; AMOP; 1.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PS50092; TSP1; 1.  
FT NON TER 1  
SQ SEQUENCE 518 AA; 58281 MW; EB6B57C4D44E928 CRC64;

Query Match 84.6%; Score 842; DB 2; Length 518;  
Best Local Similarity 98.1%; Pred. No. 2.1e-48;  
Matches 157; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 28 PRLKEEEAFLPRTHLQAEHPHQGCVTTEPAAMTGNTPPTPEVTEPLRLLEQLKP 87  
DB 1 PRLKEEEAFLPRTHLQAEHPHQGCVTTEPAAMTGNTPPTPEVTEPLRLLEQLKP 60  
QY 88 GLASTTSTPNPDQASAPDRPLREEEEARLLPRTHLQAEHPHQGCVTTEPAALTPG 147  
DB 61 GLANTTSTPNPDQASADPRPLREEEEARLLPRTHLQAEHPHQGCVTTEPAALTPG 120  
QY 148 NATPRTQEVTPLELLEQLKLPVLVHATLSTPNPDQVNTIK 187  
DB 121 NATPRTQEVTPLELLEQLKLPVLVHATLSTPNPDQVNTIK 160

## RESULT 8

Q86TW4 PRELIMINARY; PRT; 308 AA.  
AC Q86TW4;  
DT 01-JUN-2003 (TReMBLrel. 24, Created)  
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Full-length cDNA clone CS0DI024YA19 of Placenta of Homo sapiens  
DE (human) (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Li W.B., Gruber C., Jessee J., Polayes D.;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Genoscope;  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BX248280; CAD62608.1; -;  
DR InterPro; IPR005533; AMOP.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF03782; AMOP; 1.  
DR Pfam; PF00090; TSP 1; 1.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PS50092; TSP1; 1.  
FT NON TER 1  
SQ SEQUENCE 308 AA; 34735 MW; 38F47CC67F7D44E1 CRC64;

Query Match 32.3%; Score 321; DB 2; Length 308;  
Best Local Similarity 100.0%; Pred. No. 6.3e-14;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 128 AELHGHGCVTTEPAALT-PGNATPPRTQEVTPLELLEQLKLPVLVHATLSTPNP 180

QY 128 AELHGHGCVTTEPAALT-PGNATPPRTQEVTPLELLEQLKLPVLVHATLSTPNPQVNTIK 187  
DB 1 AELHGHGCVTTEPAALT-PGNATPPRTQEVTPLELLEQLKLPVLVHATLSTPNPQVNTIK 60

## RESULT 9

Q92954 PRELIMINARY; PRT; 1404 AA.  
AC Q92954;  
DT 01-FEB-1997 (TReMBLrel. 02, Created)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Megakaryocyte stimulating factor.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
RT "Purification, Biochemical Characterization, and Cloning of a Novel  
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony  
RT Stimulating Activity";  
RL Blood 78:279-279(1991).  
[2]  
RP SEQUENCE FROM N.A.  
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,  
RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,  
RA Jacobs K., Turner K.;  
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor";  
RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegenhoff J.,  
RL Mosher D.F. (eds.);  
RL BIOLOGY OF VITROTECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier  
RL Science Publishers B.V. (1993).  
[3]  
RP SEQUENCE FROM N.A.  
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U70136; AAB03089.1; -;  
DR HSSP; P04004; 10C0.  
DR Genew; HGNC:9364; PRG4.  
DR GO; GO:0008283; P:cell proliferation; TAS.  
DR InterPro; IPR000585; Hemopexin.  
DR InterPro; IPR001212; Somatomedin\_B.  
DR Pfam; PF00045; Hemopexin; 2.  
DR Pfam; PF01033; Somatomedin\_B; 2.  
DR PRINTS; PR00022; SOMATOMEDINB.  
DR SMART; SM00120; HX; 2.  
DR SMART; SM00201; SO; 2.  
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN\_1.  
DR PROSITE; PS00524; SMB 1; 2.  
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match 14.5%; Score 144; DB 2; Length 1404;  
Best Local Similarity 28.4%; Pred. No. 0.19;  
Matches 50; Conservative 11; Mismatches 97; Indels 18; Gaps 3;  
QY 23 SASPPRPLKEEEAFLPRTHLQAEHPHQGCVTTEPAAMT-----PGNTTTP 71  
DB 555 TTTKEPAPTTPEKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPKPAPTTPK 614  
QY 72 RTPVTPLELLEQLKLPGLASTTLPNPDQASAP-----DPRPLREEEEARLLPRTHLQ 127  
DB 615 ETAPTTKLTPTTPEKLAFTTPEKPAFTTPEKLAFTTPEKPAFTTPEKPAFTTPEK 674  
QY 128 AELHGHGCVTTEPAALT-PGNATPPRTQEVTPLELLEQLKLPVLVHATLSTPNP 180

```

Db 675 NTPKEPAPTPKPEAPPTPKPEAPPTPKETAPTTPKGTAPTTLKEPAPTTPKKPAP 730
RESULT 10
Q9BX49
ID Q9BX49 PRELIMINARY; PRT; 1404 AA.
AC Q9BX49;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE BGL174L6.2 (MSF: megakaryocyte stimulating factor ).
GN Name=BGL174L6.2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133553; CAC36090.1; -.
DR HSP; P04004; 1OC0.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin B; 2.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SMB_1; 2.
SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEES CRC64;

Query Match 14.5%; Score 144; DB 2; Length 1404;
Best Local Similarity 28.4%; Pred. No. 0.19; Mismatches 11; Gaps 3;
Matches 50; Conservative 11; Indels 18;

QY 23 SASDPRLKBEAEAPLLPRTHLQAEPHQHCWTVTEPAAMT-----PONTTPP 71
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 TTTKEPAPTPKPEAPPTPKKPAPTTPKEPAPTTPKEPAPTTPKPAPTTPKPA 614
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 RTPEVTPRLLEQLKPLGLASTLSTPNPDQASAP-----DPRLEBEEEARLLPRTHLQ 127
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 ETAPTPPKLTPPTPEKLAPTTPKPAPTTPPELAPTTPPEAPTTPPEAPTTPKAAAP 674
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 AELHQHCWTVTEPAALT---PGNATPRTQEVTPRLLEQLKPLGLVHATLSTNP 180
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 NTPKEPAPTPKPEAPPTPKPEAPPTPKETAPTTPKGTAPTTLKEPAPTTPKKPAP 730

RESULT 11
Q7QFG1
ID Q7QFG1 PRELIMINARY; PRT; 837 AA.
AC Q7QFG1;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE AGCP13522 (Fragment).
GN Name=agCG56847; ORFNames=ENSA0000000018269;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=PEST;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DB EMBL; AAAB01008846; EAA06308.1; -.

```

```

DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin_bind_PeRA.
DR Pfam; PF01607; CBM_14; 5.
DR PROSITE; PS00940; CHIT_BIND_II; 5.
FT NON_TER 1
FT NON_TER 837
SQ SEQUENCE 837 AA; 85371 MW; C506EF34EF2A726B CRC64;

Query Match 13.3%; Score 132; DB 2; Length 837;
Best Local Similarity 27.6%; Pred. No. 0.71;
Matches 55; Conservative 24; Mismatches 60; Indels 60; Gaps 12;

QY 11 PRPSGLTRLAERV-SASDPRLKBEAEAPLLPRTHLQAEPHQHCWTVTEPAAM-----TP 65
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 PEPSTIPLTPPGSATPKPTP-----EPSTIPLT-----TPGSATPKTPPESTIPLTP 370
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 GNTTPPTPE---VTPRLLEQLKPLGLASTLSTPNPD-----TQASASDPRLREEE 116
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 GSATPKTPPESTIPL-----TPPG-SATSEPTPEPTIPLTPPGSATPEPTP---E 420
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 EARLLPRTHLQAEHLHQHCWTVTEPAALTTCGNATPPRTQE-----VTPLLLEQLKPL 169
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 PSTNIPLT-----TPGSTTPKTPPEPTIPLTPPGSATSEVTPE 460
: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 L-VHATLSTPNPDNOVTIK 187
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 LTTNIPLTPGTASSIVE 479
: : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
PCLO RAT
ID PCLO RAT STANDARD; PRT; 5085 AA.
AC Q9JK36; Q9JL71;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN Name=Piclo;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984; DOI=10.1016/S0896-6273(00)80883-1;
RA Fenster S.D., Chung W.U., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT bassoon.";
RL Neuron 25:203-214(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CALICUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP VAL-4688; 4688-VAL-NET-4689; NET-4689; 4690-VAL-VAL-4691;
RP 4692-GLN-ASN-4693 AND ALA-4694.
RX MEDLINE=21181819; PubMed=11285225; DOI=10.1093/emboj/20.7.1605;
RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RT "An unusual C(2)-domain in the active-zone protein piccolo:
RT implications for Ca(2+) regulation of neurotransmitter release.";
RL EMBO J. 20:1605-1619(2001).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -!- SUBUNIT: Interacts with Rabac1/Prai, RIMS2 and profilin (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC junctions.
CC -!- ALTERNATIVE PRODUCTS:

```

```
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9JKS6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC -- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -- SIMILARITY: Contains 2 C2 domains.
CC -- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF138789; AAF07822.2; --
CC EMBL; AF227534; AAF63196.1; --
CC PDB; 1RH8; NMR; A=4635-4776.
CC RGD; 69406; PC10.
CC
CC GO; GO:0045202; C:synapse; IDA.
CC GO; GO:0005509; F:calcium ion binding; IDA.
CC GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
CC GO; GO:0005522; F:profilin binding; ISS.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
CC GO; GO:0016080; P:synaptic vesicle targeting; NAS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR008899; Znf_piccolo.
CC Pfam; PF00168; C2; 2.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF05715; zf-piccolo; 2.
CC SMART; SM00239; C2; 2.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00499; C2_DOMAIN_1; 1.
CC PROSITE; PS00004; C2_DOMAIN_2; 2.
CC PROSITE; PS0106; PDZ; 1.
CC 3D-structure; Alternative splicing; Calcium/phospholipid-binding;
KW Metal-binding; Repeat; Zinc; Zinc-finger.
KW DOMAIN 372 491
FT FT
FT ZN_FING 523 547
FT ZN_FING 1010 1033
FT DOMAIN 2351 2362
FT DOMAIN 4442 4536
FT DOMAIN 4653 4752
FT DOMAIN 4968 5059
FT VARSPIC 4876 4880
FT
FT VARSPIC 4881 5085
FT
FT MUTAGEN 4668 4668
FT
FT MUTAGEN 4674 4674
FT
FT MUTAGEN 4688 4688
FT
FT MUTAGEN 4688 4689
FT
FT MUTAGEN 4689 4689
FT MUTAGEN 4690 4691
FT
FT MUTAGEN 4692 4693
FT
FT MUTAGEN 4694 4694
FT
FT SEQUENCE 5085 AA; 552702 MW; SAIBS543201A7450 CRC64;
SQ
Query Match 12.5%; Score 124.5; DB 1; Length 5085;
Best Local Similarity 23.2%; Pred. No. 15;
Matches 42; Conservative 27; Mismatches 63; Indels 49; Gaps 6;
Qy 3 VKPRLGRPGSLTRLAEVSASDPRLKKEEEAPLLPRTHLQAEHQHGCWVTEPAA 62
Db 659 VKQFTLHGPTAPAPQLPVAEALPEPAPPK-EPGGL-PEQ-----A 698
Qy 63 MTPGNTTPRTPVTLRLRLKQLPLGLASTTLSTPNPDTQASASDPRLRLREEEARLLP 122
Db 699 KAPVDVEPKQPKWETRADIQ-----SSSTTKPDILSS-----QVQSOAQVKT 742
Qy 123 RTHLQAEHLHGHCWVTEPAAALTPGNATPPRTQRTQVTPLLLELKLDELVHATLSTPNPDN 182
Db 743 ASPLKTD-----SAXPSQSFPTGKTKTLPDLSKAMPSPASDKIISQPGPGS 789
Qy 183 Q 183
Db 790 E 790
RESULT 13
O17338 PRELIMINARY; PRT; 880 AA.
ID O17338
AC O17338;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein T23E7.2.
GN ORFNames=T23E7.2, T23E7.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Latreil P., Stellyes L., Elliot G., Wilson R.;
RT "The sequence of C. elegans cosmid T23E7."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF026205; AAB71258.1; -.
DR PIR; D89756; D89756.
DR WormBase; WBGene00020732; T23E7.2.
DR WormPep; T23E7.2b; CE14066.
KW Hypothetical protein.
SQ SEQUENCE 880 AA; 95398 MW; 97A8A101E8FBA1C1 CRC64;
Query Match 12.4%; Score 123; DB 2; Length 880;
Best Local Similarity 26.9%; Pred. No. 3;
Matches 45; Conservative 14; Mismatches 62; Indels 46; Gaps 7;
```



Search completed: February 18, 2005, 23:59:04  
Job time : 180 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 00:06:56 ; Search time 2780 Seconds  
(without alignments)  
3259.398 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213  
Perfect score: 995  
Sequence: 1 LPVKPLRGRPGSLTRLA.....PELVHATLSTPNPDNQVTK 187

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool\_P/US10030225/runat\_18022005\_100354\_2341/app.query.fasta\_1.327  
-DB=GenEmbl -QWMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10030225.CG1.1.5600 @runat\_18022005\_100354\_2341 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	995	100.0	2981 6 AX136231 Sequence
2	995	100.0	2981 6 BD093285 Different
3	995	100.0	2981 6 BD123568 Secretary
4	995	100.0	2981 9 AK075445 Homo sapi

5	987	99.2	2883	9	AJ583024	Homo sapi
6	964.5	96.9	3052	9	AK056709	Homo sapi
7	938.5	94.3	1698	6	CQ720015	Sequence
8	857.5	86.2	750	6	AX136531	Sequence
9	857.5	86.2	750	6	BD123771	Secretary
10	842	84.6	2783	9	BC033140	Homo sapi
c 11	841.5	84.6	231464	9	AF111168	Homo sapi
c 12	658.5	66.2	497	6	CQ049341	Sequence
c 13	658.5	66.2	497	6	CQ064365	Sequence
c 14	658.5	66.2	497	6	CQ091307	Sequence
c 15	658.5	66.2	497	6	CQ130141	Sequence
c 16	658.5	66.2	497	6	CQ168762	Sequence
c 17	658.5	66.2	497	6	CQ197880	Sequence
c 18	658.5	66.2	497	6	CQ213333	Sequence
c 19	658.5	66.2	497	6	CQ251912	Sequence
c 20	658.5	66.2	497	6	CQ289061	Sequence
c 21	658.5	66.2	497	6	CQ326064	Sequence
c 22	430	43.2	243	6	CQ054407	Sequence
c 23	430	43.2	243	6	CQ073647	Sequence
c 24	430	43.2	243	6	CQ104532	Sequence
c 25	430	43.2	243	6	CQ143244	Sequence
c 26	430	43.2	243	6	CQ178734	Sequence
c 27	430	43.2	243	6	CQ203077	Sequence
c 28	430	43.2	243	6	CQ226434	Sequence
c 29	430	43.2	243	6	CQ264575	Sequence
c 30	430	43.2	243	6	CQ301666	Sequence
c 31	430	43.2	243	6	CQ338879	Sequence
c 32	234.5	23.6	172662	10	AC120540	Mus muscu
c 33	154.5	15.5	189436	2	AL691432	Homo sapi
c 34	153.5	15.4	146664	9	AC010327	Homo sapi
c 35	150	15.1	123255	9	AC006003	Homo sapi
c 36	150	15.1	184132	2	AC079590	Homo sapi
c 37	150	15.1	200070	9	AC008746	Homo sapi
c 38	149.5	15.0	6694	6	AX252013	Sequence
c 39	149.5	15.0	6694	6	AX348784	Sequence
c 40	149	15.0	132211	9	AC091196	Homo sapi
c 41	149	15.0	170880	2	AC145870	Sequence
c 42	148	14.9	183930	2	AC141015	Rattus no
c 43	147	14.8	6456	6	AX345908	Sequence
c 44	146	14.7	132634	2	AC123294	Rattus no
c 45	146	14.7	171849	9	AC010616	Homo sapi

ALIGNMENTS

RESULT 1  
AX136231  
LOCUS AX136231 2981 bp DNA linear PAT 30-MAY-2001  
DEFINITION Sequence 153 from Patent EP1067182.  
ACCESSION AX136231  
VERSION AX136231.1 GI:14272639  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and  
Hayashi, K.  
TITLE Secretary protein or membrane protein  
JOURNAL Patent: EP 1067182-A 153 10-JAN-2001;  
FEATURES Helix Research Institute (JP)  
Location/Qualifiers  
source  
1..2981  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
58..1773  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39758.1"  
/db\_xref="GI:14272640"  
/translation="MRALRDRAGLLLCVLLAALLAALGLPVKKPLRGRPGSLTR"

LAEVSADPDRPLKEEBEAPLLPRTHLOAEPHOHCWTVTEPAAMTGNTTTPPTREV  
TEPLELQKLPGLASTTSLTPNDPTQASADPDRPLREBEERALLPRTHLOAELHOGH  
CWTVEAALTPGNATPRQTEVPTPLLELQKLPVHATLSVNPNDQVTVKVEDP  
QAEVSDLLAEPNPPQDDTLSPALWPFIMWDYKGEKDRAPGEGKEDEDEDYP  
SETEGEDQEDKEDEEALQWNGTTDNDQWLPAGDWFKDSVSYDYEPQKELRN  
WSPGSCNCSQKQRTPCGCTATRTCTDLPCPGTGTGDTGLSPSEWKLARN  
ATMDHDDVDSCEKWLKSDFLTKYLSQMLRDLPCPCAPYLEAMDSPLQDEHQG  
RSPWRDASGPREDLDIYQPTARECLRSLMSGESSTLAAOHCVCYDESRLLTRGKGAG  
MPNLISIDFSPKLFHFKFTDTPWILCKGDSRLHVLFPNNGRACDTPLEEEVLAQLQ  
EAKY"

## ORIGIN

Alignment Scores:  
Pred. No.: 1,67e-40 Length: 2981  
Score: 995.00 Matches: 187  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x AX136231 (1-2981)

Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
Db 136 CTCCCGGTGAAGAGCGCGCTCCGGGACACCGGCTGGGAGCCTCACGAGGCTCGCA 195  
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
Db 196 GAGGTCTCAGGCTCCCGAGATCTTAGGCTCTGAAGAGAGAGGAGGACCACTGCTC 255  
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
Db 256 CCCAGAACCCACCTCGAGCGAGGACCAACACATGGATGCTGACCTGTCACTGAGCCA 315  
Qy 61 AlaAlaMetThrProGlyAsnThrThrProArgThrProGluValThrProLeuArg 80  
Db 316 GCAGCCATGACCCAGGCAACACCACTCCAGAGCCCGAGAGCTTACTCCGTTGGG 375  
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
Db 376 CTGGAGCTGCAGAGCTCCCGGATGTCAGCAGCAACCTTGAGTACCCTTAACCTGAT 435  
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
Db 436 ACCCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGAGAGGAGGAGGACGACTG 495  
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
Db 496 CTCCCGAGACCCACCTCGAGCGAGGCTACACCAACATGGATGTTGACCTGTCACTGAG 555  
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProArgThrGlnGluValThrProLeu 160  
Db 556 CCAGCAGGCTTACCCCGAGGAATGCCAGCTCCAGAGCCCGAGAGGTTACTCCCTTG 615  
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
Db 616 CTGCTGGAGCTGCAGAGCTGCAGATTTGGTCCAGCAACCTTGAGTACCCCTAACCT 675  
Qy 181 AspAsnGlnValThrIleLys 187  
Db 676 GATAACCCAGGTGACCATCAAG 696

## RESULT 2

BD093285  
LOCUS BD093285 2981 bp DNA linear PAT 27-AUG-2002  
DEFINITION Differentiation growth factor.  
ACCESSION BD093285  
VERSION BD093285.1 GI:22638873  
KEYWORDS WO 0104312-A/1.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 2981)  
Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Yoshida,K. and Masuho,Y.  
TITLE Differentiation growth factor  
AUTHORS  
JOURNAL Patent: WO 0104312-A 1 18-JAN-2001;  
HELIIX RESEARCH INSTITUTE,TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,  
YURI KAWAI,KENJI YOSHIDA,YASUHIKO MASUHO  
OS Homo sapiens (human)  
COMMENT  
PN WO 0104312-A/1  
PD 18-JAN-2001  
PF 06-JUL-2000 WO 2000JP004514  
PR 08-JUL-1999 JP 99P 194179,18-OCT-1999 US 60/159586 PI  
TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,YURI KAWAI,KENJI PI  
YOSHIDA,

PI YASUHIKO MASUHO  
PC C12N15/16,C12N15/12,C12N15/85,C12N5/10,C12P21/02,C07K14/575,  
C07K14/72  
PC C12Q1/68,C12Q1/02,A61K67/027/(C12P21/02,C12R1/91) CC  
FH Key Location/Qualifiers  
FT CDS (58)..(1770).

## FEATURES

## Source

1..2981  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:  
Pred. No.: 1,67e-40 Length: 2981  
Score: 995.00 Matches: 187  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BD093285 (1-2981)

Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
Db 136 CTCCCGGTGAAGAGCGCGCTCCGGGACACCGGCTGGGAGCCTCACGAGGCTCGCA 195  
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
Db 196 GAGGTCTCAGGCTCCCGAGATCTTAGGCTCTGAAGAGAGGAGGAGGACCACTGCTC 255  
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
Db 256 CCCAGAACCCACCTCGAGCGAGGACCAACACATGGATGCTGACCTGTCACTGAGCCA 315  
Qy 61 AlaAlaMetThrProGlyAsnThrThrProArgThrProGluValThrProLeuArg 80  
Db 316 GCAGCCATGACCCAGGCAACACCACTCCAGAGCCCGAGAGGTTACTCCGTTGGG 375  
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
Db 376 CTGGAGCTGCAGAGCTCCCGGATGTCAGCAGCAACCTTGAGTACCCTTAACCTGAT 435  
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
Db 436 ACCCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGAGAGGAGGAGGACGACTG 495  
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
Db 496 CTCCCGAGACCCACCTCGAGCGAGGCTACACCAACATGGATGTTGACCTGTCACTGAG 555  
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProArgThrGlnGluValThrProLeu 160  
Db 556 CCAGCAGGCTTACCCCGAGGAATGCCAGCTCCAGAGCCCGAGAGGTTACTCCCTTG 615  
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
Db 616 CTGCTGGAGCTGCAGAGCTGCAGATTTGGTCCAGCAACCTTGAGTACCCCTAACCT 675  
Qy 181 AspAsnGlnValThrIleLys 187

## ORIGIN

Alignment Scores:  
 Pred. No.: 1,67e-40 Length: 2981  
 Score: 995.00 Matches: 187  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x AK075445 (1-2981)

Qy 1 LeuProVallylsLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 Db 136 CTCGCCGTGAAGAGCGCGCTCCGCGGACACCGGCTGGAGCCTCACGAGGCTGCA 195  
 Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 Db 196 GAGTCTCAGGCTCCCGAGTCTAGGCTCTGAGGAAGAGGAGGACCACTGCTC 255  
 Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluPro 60  
 Db 256 CCGAAGACCCACCTCGAGCGAGGACCAACATGATGCTGACTGTCACTGAGCCA 315  
 Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 Db 316 GCAGCATGACCCAGGCAACACCACTCCAGGACCCAGAGGTTTACTCCGTTGCGG 375  
 Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 Db 376 CTGGAGCTGCAGAGCTCCCGGATGCGCAGCAACCTTGAGTACCCCTTAACCTGAT 435  
 Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 Db 436 ACCAGGCTTCAGCTCCCGAGATCTAGGCTCTGAGGAGAGGAGGAGGACGACTG 495  
 Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTTPThrValThrGlu 140  
 Db 496 CTCGCCAAGACCCAGCTCGAGGAGCTACCAACATGATGTTGACTGTCACTGAG 555  
 Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 Db 556 CCAGAGCCCTGACCCCGAGGAATGCCAGCTCCAGGACCCAGGAGTTTACTCCCTTG 615  
 Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 Db 616 CTGCTGGAGCTGCAGAGCTCCAGATTGTCACCGCACTTGAGTACCCCTAACCT 675  
 Qy 181 AspAsnGlnValThrIleLys 187  
 Db 676 GATAACGAGGTGACCATCAAG 696

RESULT 5  
 AJ583024  
 LOCUS Homo sapiens mRNA for Tail1 protein. linear PRI 18-JUN-2004  
 DEFINITION  
 ACCESSION AJ583024  
 VERSION AJ583024.1 GI:48958174  
 KEYWORDS Tail1 gene; Tail1 protein; thrombospondin and AMOP containing  
 isthm-like 1 protein.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Rossi,V., Belfagna,G., Rampazzo,A., Bauce,B. and Danieli,G.A.  
 TITLE Tail1: an isthm-like gene, containing type 1  
 thrombospondin-repeat and AMOP domain, mapped to ARVD1 critical  
 region  
 JOURNAL Gene 335, 101-108 (2004)  
 PUBMED 15194193  
 REFERENCE 2 (bases 1 to 2883)  
 AUTHORS Rossi,V.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-SEP-2003) Rossi V., Department of Biology, University

# FEATURES

of Padua, viale G. Colombo, 3 - Padova, 35131, ITALY  
 Location/Qualifiers  
 1..2883  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /map="14q24.3"  
 /tissue\_type="heart"  
 /dev\_stage="adult"  
 1..2883  
 /gene="TAIL1"  
 36..1751  
 /gene="TAIL1"  
 /function="putative role in development"  
 /codon\_start=1  
 /product="thrombospondin and AMOP containing isthm-in-like  
 1 protein, (Tail1)"  
 /protein\_id="CAE47313.1"  
 /db\_xref="GI:48958175"  
 /translation="MRALRRDAGLLLCVLLAALLAALGLPVKKPRIGRRPGSLTR  
 LAEVSFDPRLPKEEEELPRLTHIQABPHQGCWTVTEPAAMTPGNATPPRIPEV  
 TPLRLEQLPLGANTTSTPNPTQASDPRLPEEEELRLPRLTHQAEUHQH  
 CWTVEPALTPGNATPPRIPEVPLWLEQLKLPVHATLSTPNPQNTIKVDEP  
 QAEVSIIDLAEPGNPPQDTLSWLPALWGLGDKYKGEKDRAPGEKEKEDEVP  
 SEDTGEDQEKDEEEQALWNGTTDNDQGLAPGDVFKDSVSDYVPOKEWSP  
 WSPGSCNCTKQORTPCGYGTATETRTCDLPSCPGTGDKDILGLPSEWKLARN  
 ATDHMDQVDSCKWLNCKSDFLIKYSQMLRDLFPCPAYPLEAMDSPVSLQEHQ  
 RSPRWDASGRPRERLDIYQTPARFLSMISGESSTLAQHCCYDEDSRLTLTGKAG  
 MPNLISDFSKLHFKFTTPWILKGDWSRLHVLPPNNGRACITDNLPLEEYLAQLQ  
 EAKY"  
 36..113  
 /sig\_peptide  
 /gene="TAIL1"  
 114..1748  
 /mat\_peptide  
 /gene="TAIL1"  
 /product="Tail1 protein"  
 2865..2870  
 /polya\_signal  
 /gene="TAIL1"

## ORIGIN

Alignment Scores:  
 Pred. No.: 4,02e-40 Length: 2883  
 Score: 987.00 Matches: 185  
 Percent Similarity: 99.47% Conservative: 1  
 Best Local Similarity: 98.93% Mismatches: 1  
 Query Match: 99.20% Indels: 0  
 DB: 9 Gaps: 0  
 US-10-030-225-2\_COPY\_27\_213 (1-187) x AJ583024 (1-2883)  
 Qy 1 LeuProVallylsLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 Db 114 CTCGCCGTGAAGAGCGCGCTCCGCGGACACCGGCTGGAGCCTCACGAGGCTGCA 173  
 Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 Db 174 GAGTCTCAGGCTCCCGAGTCTAGGCTCTGAGGAAGAGGAGGAGGACCACTGCTC 233  
 Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluPro 60  
 Db 234 CCGAAGACCCACCTCGAGGAGGACCAACATGATGCTGACTGTCACTGAGCCA 293  
 Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 Db 294 GCAGCATGACCCAGGCAACCGCCTCCAGGACCCAGAGGTTTACTCCGTTGCGG 353  
 Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 Db 354 CTGGAGCTGCAGAGCTGCCGGATTGGCCACACACCTTGAGTACCCCTTAACCTGAT 413  
 Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 Db 414 ACCAGGCTTCAGGCTCCCGAGATCTTAGGCTCTGAGGGAAGAGGAGGAGGACGACTG 473









Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.L., Skalska, U., Smalilus, D.E., Schnercher, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 2783)  
Strausberg, R.  
Direct Submission  
Submitted (25-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu, Parvaneh Saeedi, JR Santos, Angelique Schnercher, Ursula Skalska, Duane Smalilus, Jeff Stott, Miranda Tsai, George Yang, Jacque Schein, Asim Siddiqui, Rob Holt, Marco Marra.

```

repeat_region      /rpt_family="L1MD1"
200. .494
/rpt_family="AluSx"
repeat_region      500. .1283
/rpt_family="AluSg"
repeat_region      1284. .1581
/rpt_family="AluSg"
repeat_region      1582. .1774
/rpt_family="AluJb"
repeat_region      1775. .2058
/rpt_family="AluJb"
repeat_region      2059. .2934
/rpt_family="L191_Send"
repeat_region      complement(2937. .3255)
/rpt_family="MER4A"
repeat_region      3323. .3687
/rpt_family="L1"
repeat_region      3688. .3977
/rpt_family="AluSg"
repeat_region      4009. .4314
/rpt_family="AluJo"
repeat_region      4321. .4934
/rpt_family="L1"
repeat_region      4940. .5236
/rpt_family="AluSp"
repeat_region      5238. .5349
/rpt_family="L1MC1"
repeat_region      5350. .5650
/rpt_family="AluSq"
repeat_region      5888. .5924
/rpt_family="TAAAA)n"
repeat_region      5869. .5940
/rpt_family="CAAAA)n"
repeat_region      6086. .6271
/rpt_family="AluJb"
repeat_region      complement(6500. .6588)
/rpt_family="5S"
unsure              complement(6715. .6815)
                    /note="low quality data"
repeat_region      complement(6914. .7458)
/rpt_family="LINE2"
repeat_region      7523. .7707
/rpt_family="L1MB6"
repeat_region      7708. .8007
/rpt_family="AluSx"
repeat_region      8233. .8305
/rpt_family="TA)n"
repeat_region      9186. .9230
/rpt_family="CA)n"
CDS                 complement(join(9248. .9470,10060. .10240,16529. .16722,
25330. .25484))
/note="Intron-exon boundaries defined in relation to an
EST contigs that includes AA397580, A1138314, AA399545,
AA68184, all from testis. The closest match in BLASTX is
to E. coli pyruvate-5-carboxylate reductase."
/codon_start=1
/product="unknown"
/protein_id="AAD09625.1"
/db_xref="GI:4186188"
/translation="MDMLQDLESIQFYGVPEEDRIWLYIQGRSGLMIEACAHATFF
CKLYNLRASLNKNQSRHLSIGLSNATPEEPKVGILGGHGLGKQLAGTLQIGTIP
AESLRITRRPETGELQKIKCFYHNADLVSWADVIFLCCLPSQLPNICVETYSL
EKASIVSPVAALPLPKLLLNHTNLRQYQYDEDSVSWGANKGVIAALQDPTIL
QATCPYPAGHGTGYFVHSPENYFPVGPLLSY"
repeat_region      complement(9631. .9708)
/rpt_family="LINE2"
repeat_region      complement(10461. .10762)
/rpt_family="AluSx"
repeat_region      complement(10804. .11077)
/rpt_family="AluJb"
repeat_region      complement(11156. .11306)
/rpt_family="MIR"
repeat_region      complement(11493. .11615)
/rpt_family="L1MB5"
repeat_region      /rpt_family="FLAM_C"
11800. .12161
/rpt_family="THEIC"
complement(12406. .12695)
/rpt_family="AluSx"
12746. .12920
/rpt_family="BC200"
12921. .13217
/rpt_family="AluSx"
13236. .13751
/rpt_family="L1MB1"
13783. .14084
/rpt_family="AluSx"
14085. .14403
/rpt_family="L1MB2"
14405. .14717
/rpt_family="AluSx"
14795. .14857
/rpt_family="L1MC2"
complement(14884. .14905)
/rpt_family="AT rich"
complement(15106. .15393)
/rpt_family="AluSx"
complement(15606. .15818)
/rpt_family="L1MB2"
complement(15835. .16131)
/rpt_family="AluSc"
complement(16775. .17073)
/rpt_family="AluY"
17491. .17777
/rpt_family="AluJo"
17778. .17804
/rpt_family="(CA)n"
complement(17815. .17882)
/rpt_family="L1MB2"
17883. .18175
/rpt_family="AluSx"
18177. .18251
/rpt_family="(TAGA)n"
complement(18244. .18306)
/rpt_family="L1MB1"
complement(18310. .18609)
/rpt_family="AluSx"
complement(18624. .18771)
/rpt_family="L1MB3"
complement(18765. .18847)
/rpt_family="L1"
18858. .19501
/rpt_family="SVA"
complement(19503. .20417)
/rpt_family="L1"
complement(20104. .20625)
/rpt_family="L143_Send"
complement(20724. .21029)
/rpt_family="AluJb"
complement(21030. .21122)
/rpt_family="(TAAA)n"
complement(21125. .21423)
/rpt_family="AluJo"
complement(21486. .21662)
/rpt_family="L191_Send"
complement(21664. .21966)
/rpt_family="AluSx"
complement(22042. .22342)
/rpt_family="AluY"
complement(22350. .22937)
/rpt_family="L143_Send"
complement(22975. .23090)
/rpt_family="FLAM_C"
complement(23164. .23287)
/rpt_family="L143_Send"
complement(26429. .26704)
/rpt_family="L1MB5"

```





Job time : 2828 secs

Search completed: February 22, 2005, 02:21:04

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 00:04:01 ; Search time 423 Seconds

(without alignments)

2617.001 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKPLRGPRGSLRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table:

BLOSUM62  
-Q=/cgn2 1/USPTO spool p/US10030225/runat 18022005 100354 2333/app query.fasta\_1.327  
-DB=N Geneseq 16Dec04 -PMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOPCIL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US10030225 @CGN 1 1 708 @runat 18022005 100354 2333 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2 1/USPTO spool p/US10030225/runat 18022005 100354 2333/app query.fasta\_1.327  
-DB=N Geneseq 16Dec04 -PMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOPCIL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US10030225 @CGN 1 1 708 @runat 18022005 100354 2333 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 16Dec04:\*

1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001bs: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002as: \*  
7: Geneseqn2002bs: \*  
8: Geneseqn2003as: \*  
9: Geneseqn2003bs: \*  
10: Geneseqn2003cs: \*  
11: Geneseqn2003ds: \*  
12: Geneseqn2004as: \*  
13: Geneseqn2004bs: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	995	100.0	2981	4	AAF29348 Human pro
2	995	100.0	2981	5	AAF93820 Human CDN
3	857.5	86.2	750	5	AAF94019 Primer sp
C 4	658.5	66.2	497	4	AAI10232 Probe #16
C 5	658.5	66.2	497	4	ABA51868 Human foe

C 6	658.5	66.2	497	4	AAI31480
C 7	658.5	66.2	497	4	ABA21690
C 8	658.5	66.2	497	4	AAK25606
C 9	658.5	66.2	497	4	AAK00167
C 10	658.5	66.2	497	4	ABs25182 Human liv
C 11	658.5	66.2	497	5	AAI00170
C 12	658.5	66.2	497	6	ABs00175 Human gen
C 13	430	43.2	243	4	AAI19514
C 14	430	43.2	243	4	ABA64531
C 15	430	43.2	243	4	AAI44705
C 16	430	43.2	243	4	ABA46659
C 17	430	43.2	243	4	ABA31664
C 18	430	43.2	243	4	AAK38709
C 19	430	43.2	243	4	AAK12982
C 20	430	43.2	243	4	ABs38283
C 21	430	43.2	243	5	AAI05236
C 22	430	43.2	243	6	ABs12780
C 23	149.5	15.0	6694	6	ABL70352
C 24	149.5	15.0	6694	6	AAK61313
C 25	147	14.8	6456	6	ABL33006
C 26	145.5	14.6	74787	12	ADQ74672
C 27	144	14.5	3813	10	ADL13808
C 28	144	14.5	3936	10	ADL13807
C 29	144	14.5	4092	10	ADL13806
C 30	144	14.5	4140	12	ADK67916
C 31	144	14.5	4215	10	ADL13804
C 32	144	14.5	4215	10	ADL13805
C 33	144	14.5	4289	12	ADK67917
C 34	144	14.5	4575	8	ABX63303
C 35	144	14.5	4847	10	ADK65818
C 36	144	14.5	5008	2	AAQ27223
C 37	144	14.5	5012	10	ADL13810
C 38	144	14.5	5041	8	ABZ34656
C 39	144	14.5	5041	8	ABZ34732
C 40	144	14.5	5041	8	ABZ34731
C 41	144	14.5	5041	8	ACA64826
C 42	144	14.5	5041	10	ADD12655
C 43	144	14.5	5041	10	ADL13803
C 44	144	14.5	5041	11	ADL131655
C 45	144	14.5	5041	12	ADM98015

## ALIGNMENTS

RESULT 1

AAF29348

ID AAF29348 standard; cDNA; 2981 BP.

XX AAF29348;

AC AAF29348;

XX AAF29348;

DT 20-APR-2001 (first entry)

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

XX AAF29348;

Human proliferation differentiation factor cDNA sequence.

Human, proliferation differentiation factor; haematopoietic function; ss.

Homo sapiens.

WO200104312-A1.

18-JAN-2001.

06-JUL-2000; 2000WO-JP004514.

08-JUL-1999; 99JP-00194179.

18-OCT-1999; 99US-0159586P.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;

WPI; 2001-138354/14.

P-PSDB; AAB49765.





```
Db 196 GAGGTCTCAGGCTCCCGAGATCTTAGGCTCTGAAGGAGAGGAGGACCACTGCTC 255
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 256 CCCAGACCCACCTGCAGGAGAGCCACCAACATGGATGCTGGACTGTCTACTGAGCCA 315
Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 316 GCAGCCATGAGCCAGGAGCAACACCCCTCCAGGACCCAGAGGTTTACTCCGTTGGG 375
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100
Db 376 CTGGAGCTGCAGAGCTCCGGGATGGCCAGCACCACTTGATGACCCCTAAACCTGAT 435
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db 436 ACCCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGAGAGGAGGAGGAGGAGGAG 495
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
Db 496 CTCCCGAGACCCACCTGCAGGAGAGCTACCAACATGGATGTTGACTGTCTCACTGAG 555
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
Db 556 CCAGAGCCCTGACCCCGAGGAATGCCACGCTCCAGGACCCAGGAGGTTACTCCCTTG 615
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
Db 616 CTGTGGAGCTGCAGAGCTGCCAGATTGGTCCAGCAACCTTGAGTACCCCTAACCT 675
Qy 181 AspAsnGlnValThrIleLys 187
Db 676 GATAACCAAGTGACCATCAAG 696

RESULT 3
AAF94019 standard; DNA; 750 BP.
XX
AC AAF94019;
DT 23-MAY-2001 (first entry)
XX
DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 453.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
XX rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS Synthetic.
XX
EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-00114090.
XX
PR 08-JUL-1999; 99JP-00194179.
PR 11-JAN-2000; 2000JP-00118775.
PR 02-MAY-2000; 2000JP-00183766.
XX
PA (HELI-) HELIX RES INST.
XX
Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX WPI; 2001-093989/11.
XX
Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX gene therapy or as candidate target molecules in drug development.
XX
PS Claim 4; SEQ ID NO 453; 609pp + Sequence Listing; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by AAB88317
CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
```

```
CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
CC invention. The invention also includes methods for the production of
CC antibodies directed against the proteins, and cDNA sequences, which can
CC be used in vaccines. The polynucleotide sequences can be used in gene
CC therapy. The polynucleotide sequences and the proteins they encode may be
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate secretory protein/membrane protein expression. The
CC nucleic acids and complementary sequences may also be used as DNA probes
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
CC and quantitate the presence of similar nucleic acid sequences in samples.
CC They may also be used to study the expression and function of secretory
CC proteins/membrane polypeptides and their role in metabolism. The
CC polypeptides may be used as antigens in the production of antibodies
CC against them and in assays to identify modulators (agonists and
CC antagonists) of expression and activity. The antibodies and antagonists
CC may also be used as therapeutic agents to down regulate expression and
CC activity. The antibodies may also be used as diagnostic agents for
CC detecting the presence of the polypeptides in samples (e.g. by enzyme
CC linked immunosorbant assay (ELISA). Examples of diseases which may be
CC treated include rheumatoid arthritis and diabetes
XX
SQ Sequence 750 BP; 170 A; 249 C; 215 G; 109 T; 0 U; 7 Other;
```

```
Alignment Scores:
Pred. NO.: 8.67e-43 Length: 750
Score: 857.50 Matches: 175
Percent Similarity: 93.65% Conservative: 2
Best Local Similarity: 92.59% Mismatches: 6
Query Match: 86.18% Indels: 6
DB: 5 Gaps: 1
```

US-10-030-225-2\_COPY\_27\_213 (1-187) x AAF94019 (1-750)

```
Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db 136 CTCCCGTGAAGAGCGCGGCTCCGGAGCACCGCTGGAGCCTCAGGAGGCTCGCA 195
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db 196 GAGGTCTCAGCTCCCGAGATCTTAGGCTCTGAAGGAGAGGAGGAGGAGGAGGAGGAG 255
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 256 CCCAGACCCACCTGCAGGAGGAGCCACCAACATGGATGCTGACTGTCTCACTGAGCCA 315
Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 316 GCAGCCATGAGCCAGGAGCAACACCCCTCCAGGACCCAGAGGTTTACTCCGTTGGG 375
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAla-SerThrThrLeuSerThrProAsnProAs 100
Db 376 CTGGAGCTGCAGAGCTCCGGGATGGCCAGCACCAACCTTGATGACCCCTAAACCTGA 435
Qy 100 pThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db 436 TACCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGAGAGGAGGAGGAGGAGGAGGAG 495
Qy 120 uLeuProArgThrHisLeuGlnAla-GluLeuHisGlnHisGlyCysTrpThrValThrG 140
Db 496 GCTCCCGAGACCCACCTGCAGGAGAGCTTACCAACATGGATGTTGACTGTCTCACTG 555
Qy 140 luProAlaAlaLeuThrProGlyAsnAlaThrPro-ProArgThrGlnGluValThrPro 159
Db 556 AGCCAGAGCCCTGACCCAGGGAATGCCAGCTCCAGGACCCAGAGGAGGAGGAGGAGGAGGAG 615
Qy 160 LeuLeuLeuGluLeuGlnLys---LeuProGluLeuValHisAla-ThrLeuSer-ThrP 178
Db 616 TTGCTGTGGAGCTGCAAGAAAGCTGCCAGAAATTGGTCCAGCAACCTTGGAGTTANCC 675
Qy 178 roAsnProAspAsnGln 183
Db 676 CTAAACCTGATAACCA 692
```

XX	AAI10232 standard; DNA; 497 BP.
XX	AAI10232;
XX	12-OCT-2001 (first entry)
DE	Probe #165 for gene expression analysis in human cervical cell sample.
XX	Probe; human; microarray; gene expression; cervical epithelial cell;
XX	cervical cancer; ss.
OS	Homo sapiens.
PN	WO200157278-A2.
PD	09-AUG-2001.
XX	30-JAN-2001; 2001WO-US000670.
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	(MOLE-) MOLECULAR DYNAMICS INC.
PA	Penn SG, Hanzel DK, Chen W, Rank DR;
PI	WPI; 2001-488901/53.
PT	Human genome-derived single exon nucleic acid probes useful for analyzing
PT	gene expression in human cervical epithelial cells.
XX	Claim 25; SEQ ID NO 165; 487bp; English.
XX	The present invention relates to human single exon nucleic acid probes
CC	(SENP). The present sequence is one such probe. The SENPs are derived
CC	from human HeLa cells. The SENPs can be used to produce a single exon
CC	microarray, which can be used for measuring human gene expression in a
CC	sample derived from human cervical epithelial cells. By measuring gene
CC	expression, the probes are therefore useful in grading and/or staging of
CC	diseases of the cervix, notably cervical cancer. Note: The sequence data
CC	for this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 497 BP; 81 A; 128 C; 177 G; 111 T; 0 U; 0 Other;
SQ	
Alignment Scores:	
Pred. No.:	4,11e-31 Length: 497
Score:	658.50 Matches: 126
Percent Similarity:	76.97% Conservative: 1
Best Local Similarity:	76.36% Mismatches: 1
Query Match:	66.18% Indels: 37
DB:	Gaps: 1
US-10-030-225-2_COPY_27_213 (1-187) x AAI10232 (1-497)	
Qy	32 LysGluGluGluAlaProLeuLeuProArgThrHisLeuGlnAlaGluProHisGln 51
Db	496 AAGAGAGAGAGAGAGCACCATTGTCCCCAGAACCCACTTCGAGGCAGGCCACCAA 437
Qy	52 HisGlyCysTrpThrValThrGluProAlaMetThrProGlyAsnThrProPro 71
Db	436 CATGGATGCTGGACTGTCACTAGCCAGGCATGACCCCAGGCAAGCGCACCCCTCCC 377
Qy	72 ArgThrProGluValThrProLeuArgGluLeuGlnLysLeuProGlyLysAlaSer 91

RESIT.T 7



ABS25182/C

PR 03-AUG-2000; 2000US-00632366.

```

ID XX ABS25182 standard; DNA; 497 BP.
AC XX
XX XX
DT XX
DE XX
DE XX
XX XX
XX XX
KW Human liver single exon probe, SEQ ID NO 172.
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX XX
OS Homo sapiens.
XX XX
PN WO200157273-A2.
XX XX
PD 09-AUG-2001.
XX XX
PF 30-JAN-2001; 2001WO-US000664.
XX XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
DR WPI; 2001-488898/53.
XX XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX XX
PS Claim 1; SEQ ID NO 172; 658pp; English.
XX XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 1309 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcr_sequences
XX XX
SQ Sequence 497 BP; 81 A; 128 C; 177 G; 111 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4,11e-31 Length: 497
Score: 658.50 Matches: 126
Percent Similarity: 76.97% Conservative: 1
Best Local Similarity: 76.36% Mismatches: 1
Query Match: 66.18% Indels: 37
DB: 4 Gaps: 1

US-10-030-225-2_COPY_27_213 (1-187) x ABS25182 (1-497)

QY 32 LysGluGluGluGluAlaProLeuLeuProArgThrHisLeuGlnAlaGluProHisGln 51
Db 496 AAGGAAGAGGAGGAGCCACCTGCTCCCGAAGCCACCTGCAGGAGGAGCCACACCAA 437
QY 52 HisGlyCysTrpThrValThrGluProAlaAlaMetThrProGlyAsnThrProPro 71
Db 436 CATGATGCTGGACTGTCTACTGAGCCAGCAGCATGACCCAGGCAAGCCACCCCTCCC 377

QY 72 ArgThrProGluValThrProLeuArgLeuGluLeuGlnLysLeuProGlyLeuAlaSer 91
Db 376 AGGACCCCGAGAGGTACTCCGTTGGGCTGGAGCTGCAGAGAGCTGCCGGGATGGCCAA 317
QY 92 ThrThrLeuSerThrProAsnProAspThr----- 101
Db 316 ACAACCTTGAGTACCCCTAACCCCTGATACCCAGGTGAGAGCTACAGAGGGCCAGCAGCT 257
QY 101 ----- 101
Db 256 CCTGGCTCTGCTGGCCAGCCTTTTCTGGGTCCCTGTTGGTAAGCCCTAACCTGTCACC 197
QY 102 -----GlnAlaSerProAspProArgProLeuArgGlu 114
Db 196 TCACCTTCCCTCTCTCTTCTCAGGCTTCAGCCTCCCGAGATCCTAGGCTCTGAGGGA 137
QY 115 GluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGly 134
Db 136 GAGGAGGAGGACGACTGCTCCCGAAGCCACCTGCAGGAGAGCTACACCAACATGGA 77
QY 135 CysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArgThr 154
Db 76 TGTGGACTGTCTACTGAGCCAGCAGCCTGACCCAGGGAATGCCACGCTCCCGAGGACC 17
QY 155 GlnGluValThrPro 159
Db 16 CAGGAGGTACTCCC 2

RESULT 11
AAI00170/c
ID AAI00170 standard; DNA; 497 BP.
XX XX
AC AAI00170;
XX XX
DT 09-OCT-2001 (first entry)
XX XX
DE Probe #161 used to measure gene expression in human breast sample.
XX XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX XX
OS Homo sapiens.
XX XX
PN WO200157270-A2.
XX XX
PD 09-AUG-2001.
XX XX
PF 29-JAN-2001; 2001WO-US000661.
XX XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
DR WPI; 2001-476286/51.
XX XX
PT Novel single exon nucleic acid probe used to measuring gene expression in
PT a human breast.
XX XX
PS Claim 25; SEQ ID NO 161; 322pp; English.
XX XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
```



```
Pred. No.: 4.11e-31 Length: 497
Score: 658.50 Matches: 126
Percent Similarity: 76.97% Conservative: 1
Best Local Similarity: 76.36% Mismatches: 1
Query Match: 66.18% Indels: 37
DB: 6 Gaps: 1

US-10-030-225-2_COPY_27_213 (1-187) x ABS00175 (1-497)

Qy 32 LysGluGluGluAlaProLeuLeuProArgThrHisLeuGlnAlaGluProHisGln 51
Db 496 AAGGAAGAGGAGGAGCCACTGCTCCCAAGAACCCCTGCAGGAGAGCCACACAA 437
Qy 52 HisGlyCysTTPThrValThrGluProAlaAlaMetThrProGlyAsnThrThrPro 71
Db 436 CATGATGCTGGAGTCTCACTGAGCCAGCAGCCATGACCCAGGCAAGCCACCCCTCC 377
Qy 72 ArgThrProGluValThrProLeuArgLeuGluGlnLysLeuProGlyLeuAlaSer 91
Db 376 AGGACCCAGAGGTTACTCCGTTGGCTGGAGCTGCAGAAAGCTGCCGGGATTGGCCAAC 317
Qy 92 ThrThrLeuSerThrProAsnProAspThr----- 101
Db 316 ACAACCTTGAGTACCCTTAACCTGATACCCAGGTGAGAGCTACAGAGGGCCAGAGCT 257
Qy 101 ----- 101
Db 256 CTGGCTCTGCTGGCCAGCCCTTTCTGGGTCCCTGTTGGTAAGCCCTAACCTGTACCC 197
Qy 102 -----GlnAlaSerAlaSerProAspProArgProLeuArgGlu 114
Db 196 TCACCTTCCCTCTCTCTCTCAGGCTTCAGCTCCCCAGATCCTTAGGCTCTGAGGGA 137
Qy 115 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGly 134
Db 136 GAGGAGGAGGACGACTGCTCCCAAGAACCCACCTGCAGGAGAGAGCTACACCAATGGA 77
Qy 135 CysTTPThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArgThr 154
Db 76 TGTGGACTGTCTGACGACGAGCCCTGACCCAGGGANTGCCACCCCTCCAGGACC 17
Qy 155 GlnGluValThrPro 159
Db 16 CAGGAGGTTACTCCC 2

RESULT 13
AA119514/c
ID AA119514 standard; DNA; 243 BP.
AC AA119514;
XX
DT 12-OCT-2001 (first entry)
DE
DE Probe #9447 for gene expression analysis in human cervical cell sample.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000670.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
```

```
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
PS Claim 25; SEQ ID NO 9447; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 243 BP; 34 A; 61 C; 89 G; 59 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.28e-18 Length: 243
Score: 430.00 Matches: 79
Percent Similarity: 98.77% Conservative: 1
Best Local Similarity: 97.53% Mismatches: 1
Query Match: 43.22% Indels: 0
DB: 4 Gaps: 0

US-10-030-225-2_COPY_27_213 (1-187) x AA119514 (1-243)

Qy 22 ValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeuPro 41
Db 243 GTCTCAGCCTCCCCAGATCCTTAGGCTCTGAAGGAGAGGAGGAGGAGGAGGAGGAGG 184
Qy 42 ArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluProAla 61
Db 183 AGAACCCACCTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 124
Qy 62 AlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeu 81
Db 123 GCCATGACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 64
Qy 82 GluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrThrProAsnProAspThr 101
Db 63 GAGCTGCAGAGAGTCCCGGATTGGCCCAACACCAACCTTGAGTACCCCTAACCCCTGATACC 4
Qy 102 Gln 102
Db 3 CAG 1

RESULT 14
ABA64531/c
ID ABA64531 standard; DNA; 243 BP.
XX
XX ABA64531;
AC
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #12836.
DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
```



[illegible]

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 01:18:16 ; Search time 151 Seconds  
(without alignments)  
2026.382 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995  
Sequence: 1 LPVKPLRGRPGSUTRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool\_p/US10030225/runat\_18022005\_100355\_2367/app.query.fasta\_1.327  
-DB=Issued\_Patents\_NA -QPM=fastap -SUFFIX=nni -MINMATCH=0.1 -LOOPCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10030225 @CGN 1 1 105 /runat\_18022005\_100355\_2367 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	15.1	767677	4	US-09-949-016-12147
2	150	15.1	767677	4	Sequence 12147, A
3	144	14.5	2824	3	US-09-949-016-17361
4	144	14.5	3066	3	Sequence 13, Appl
5	144	14.5	3117	3	Sequence 83, Appl
6	144	14.5	3148	3	Sequence 73, Appl
7	144	14.5	3420	3	Sequence 57, Appl
8	144	14.5	3813	3	Sequence 103, Appl
9	144	14.5	3936	3	Sequence 43, Appl
10	144	14.5	3942	3	Sequence 41, Appl
11	144	14.5	3945	3	Sequence 141, Appl
12	144	14.5	3963	3	Sequence 49, Appl
					Sequence 45, Appl

13	144	14.5	3963	3	US-07-757-022B-59	Sequence 59, Appl
14	144	14.5	4065	3	US-07-757-022B-47	Sequence 47, Appl
15	144	14.5	4086	3	US-07-757-022B-39	Sequence 39, Appl
16	144	14.5	4092	3	US-07-757-022B-51	Sequence 51, Appl
17	144	14.5	4215	3	US-07-757-022B-61	Sequence 61, Appl
18	144	14.5	4847	4	US-10-164-595-57	Sequence 57, Appl
19	144	14.5	5008	3	US-07-757-022B-1	Sequence 1, Appl
20	144	14.5	5017	4	US-09-949-016-4956	Sequence 4956, Ap
21	144	14.5	5041	4	US-09-023-655-981	Sequence 2, Appl
22	144	14.5	5041	4	US-09-298-970A-2	Sequence 981, App
23	144	14.5	21707	4	US-09-949-016-16698	Sequence 16698, A
24	138.5	13.9	38239	4	US-09-949-016-12348	Sequence 12348, A
25	138.5	13.9	38252	4	US-09-949-016-13570	Sequence 13570, A
26	136.5	13.7	59853	4	US-09-949-016-13618	Sequence 13618, A
27	136.5	13.7	59853	4	US-09-949-016-13619	Sequence 13619, A
28	136.5	13.7	59853	4	US-09-949-016-13620	Sequence 13620, A
29	136.5	13.7	59853	4	US-09-949-016-13621	Sequence 13621, A
30	136.5	13.7	59853	4	US-09-949-016-13622	Sequence 13622, A
31	136.5	13.7	59853	4	US-09-949-016-13623	Sequence 13623, A
32	136.5	13.7	59853	4	US-09-949-016-13624	Sequence 13624, A
33	136.5	13.7	59853	4	US-09-949-016-13625	Sequence 13625, A
34	135	13.6	18722	4	US-09-949-016-15892	Sequence 15892, A
35	133	13.4	15695	4	US-09-949-016-15644	Sequence 15644, A
36	132	13.3	2130	4	US-09-909-962A-7	Sequence 7, Appl
37	131.5	13.2	80161	3	US-09-036-987A-1	Sequence 1, Appl
38	131.5	13.2	80161	3	US-09-370-700-1	Sequence 1, Appl
39	131.5	13.2	80161	4	US-09-603-207-1	Sequence 1, Appl
40	130	13.1	14462	4	US-09-902-540-1090	Sequence 1090, Ap
41	130	13.1	14462	4	US-09-902-540-9597	Sequence 9597, Ap
42	127.5	12.8	601	4	US-09-949-016-198377	Sequence 198377, A
43	127.5	12.8	92344	4	US-09-949-016-16802	Sequence 16802, A
44	127	12.8	27630	4	US-09-949-016-16362	Sequence 16362, A
45	126.5	12.7	1131	6	5168049-2	Patent No. 5168049

ALIGNMENTS

RESULT 1

US-09-949-016-12147  
; Sequence 12147, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12147  
; LENGTH: 767677  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(767677)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12147

Alignment Scores:  
Pred. No.: 4.93  
Score: 150.00  
Percent Similarity: 42.00%  
Best Local Similarity: 32.00%  
Query Match: 15.08%  
DB: 4  
Length: 767677  
Matches: 64  
Conservative: 20  
Mismatch: 63  
Indels: 53  
Gaps: 11



APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseerr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2824 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2823  
US-07-757-022B-13

Alignment Scores:  
Pred. No.: 0.00919 Length: 2824  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-13 (1-2824)

Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1307 CCACCCCTGAGAGCCCGCCACCACCCCTGAG-GAGCTCGCACCACCACCCCTGAG 1365  
Qy 26 ProArgProArgProLeuLeuGluGluAlaProLeuLeuProArgThrHisLeu 45  
Db 1366 ---GAGCCACACCCACCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGAGCGGCT 1422  
Qy 46 GluAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1423 CCCAACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1482  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1483 AAGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1530  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1531 -----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCCACTACCTCC 1578  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1579 AAGAGCTGCTCCCAAGAGCTGCTCCACCCACCACCAAGGAGCCCACTCCACCAACC 1638  
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1639 TCTGACAGCCGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACCCCTAAGGAGCT 1698  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153

Db 1699 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACC 1758  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1759 CTCAAGGAACCTGCACCCCACTACTCCCAAGAGCTGCTCCCAAGGAGCTTGCACCCACC 1818  
Qy 174 ThrLeuSerThrPro 178  
Db 1819 ACCACCAAGGGGCC 1833

## RESULT 4

US-07-757-022B-83  
Sequence 83, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/07/757,022B  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseerr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3066 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3066  
US-07-757-022B-83

Alignment Scores:  
Pred. No.: 0.0103 Length: 3066  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12

Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-83 (1-3066)

QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
DB 1550 CCACCCCTGAGAACCCGACCCACCCCTGAG-GAGCTCGACCCACCCACCCCTGAG 1608  
QY 26 ProArgProArgProLeuGluGluAlaProLeuLeuProArgThrHisLeu 45  
DB 1609 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGACGGCT 1665  
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
DB 1666 CCCAACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1725  
QY 65 -----ProGlyAsnThrProArgThrProGluValThrProLeuArgLeuGlu 82  
DB 1726 AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1773  
QY 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
DB 1774 -----AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1821  
QY 103 AlaSerAlaSerPro-----AppProArgProLeuArg 113  
DB 1822 AAGAAGCTGCTCCCAAGGAGCTTGCACCCACCCACCCACCAAGGAGCCACATCCACACC 1881  
QY 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
DB 1882 TCTGACAAAGCCGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCT 1941  
QY 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153  
DB 1942 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCC 2001  
QY 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
DB 2002 CTCAGGAACCTGCACCCACTACTCCCAAGAGCTGCCCCCAAGGAGCTGTCACCCACC 2061  
QY 174 ThrLeuSerThrPro 178  
DB 2062 ACCACCAAGGGGCC 2076

RESULT 5  
US-07-757-022B-73  
Sequence 73, Application US/07757022B  
Patent No. 643142  
GENERAL INFORMATION:  
APPLICANT: Geener, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserik, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 73:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3117 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1...3114  
US-07-757-022B-73

Alignment Scores:  
Pred. No.: 0.0105 Length: 3117  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-73 (1-3117)

QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
DB 1502 CCACCCCTGAGAACCCGACCCACCCCTGAG-GAGCTCGACCCACCCACCCCTGAG 1560  
QY 26 ProArgProArgProLeuGluGluAlaProLeuLeuProArgThrHisLeu 45  
DB 1561 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGACGGCT 1617  
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
DB 1618 CCCAACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1677  
QY 65 -----ProGlyAsnThrProArgThrProGluValThrProLeuArgLeuGlu 82  
DB 1678 AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1725  
QY 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
DB 1726 -----AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1773  
QY 103 AlaSerAlaSerPro-----AppProArgProLeuArg 113  
DB 1774 AAGAAGCTGCTCCCAAGGAGCTTGCACCCACCCACCCACCAAGGAGCCACATCCACACC 1833  
QY 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
DB 1834 TCTGACAAAGCCGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCT 1893  
QY 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153  
DB 1894 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACC 1953

Qy 154 ThrGlnGluValThrProLeuLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1954 CTCAAGGAACCTGCACCACTACTCCCAAGAGCTTGCACCCACC 2013  
Qy 174 ThrLeuSerThrPro 178  
Db 2014 ACCACCAAGGGGCC 2028

## RESULT 6

US-07-757-022B-57  
; Sequence 57, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3148 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3147

US-07-757-022B-57  
Alignment Scores:  
Pred. No.: 0.0106  
Score: 144.00  
Percent Similarity: 35.14%  
Best Local Similarity: 28.65%  
Query Match: 14.47%  
DB: 3  
Length: 3148  
Matches: 53  
Conservative: 12  
Mismatch: 99  
Indels: 22  
Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-57 (1-3148)  
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1631 CCACCCCTGAGAGCCCGCCACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1689  
Qy 26 ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45  
Db 1690 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGAGCGGCT 1746  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1747 CCCAACACCCCTAAGGAGCTGCTCCCACTACCCCTAAGGAGCTGCTCCCACTACCCCT 1806  
Qy 65 -----ProGlyAenThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1807 AAGGAGCTGCTCCCACTACCCCTAAGGAGCTGCTCCCACTACCCCT----- 1854  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrLeuSerThrProAenProAspThrGln 102  
Db 1855 -----AAAGGGAGCTGCTCCCACTACCCCTCAAGGAACCTGCACCCACTACTCCC 1902  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1903 AAGAGCTGCCCCCAAGGAGCTTGCACCCACCACTCAAGGAGCCACATCCACCACC 1962  
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1963 TCTGACAAAGCCCGCTCCCACTACCCCTAAGGGAGCTGCTCCCACTACCCCTAAGGAGCT 2022  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAenAlaThrProProArg 153  
Db 2023 GCTCCAACTACCCCTAAGGAGCTGCTCCCACTACCCCTAAGGGAGCTGCTCCCACTACC 2082  
Qy 154 ThrGlnGluValThrProLeuLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2083 CTCAAGGAACCTGCACCCACTACTCCCAAGAGCTGCCCCCAAGGAGCTTGCACCCACC 2142  
Qy 174 ThrLeuSerThrPro 178  
Db 2143 ACCACCAAGGGGCC 2157

## RESULT 7

US-07-757-022B-103  
; Sequence 103, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991

```
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/546,114
;; FILING DATE: 29-JUN-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/457,196
;; FILING DATE: 29-DEC-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/390,901
;; FILING DATE: 08-AUG-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cseri, Luann
;; REGISTRATION NUMBER: 31,822
;; REFERENCE/DOCKET NUMBER: GI 5190
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617)876-1170
;; TELEFAX: (617)876-5851
;; INFORMATION FOR SEQ ID NO: 103:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3420 base pairs
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..3420
US-07-757-022B-103
Alignment Scores:
Pred. No.: 0.0119 Length: 3420
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
Gaps: 4
US-10-030-225-2_COPY_27_213 (1-187) x US-07-757-022B-103 (1-3420)
QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1904 CCACCCCTGAGAAGCCGCGCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1962
QY 26 ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1963 ---GAGCCACACCCACACCCCTGAGGAGCTGCTCCACCACTCCCAAGCAGCGCT 2019
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
Db 2020 CCACACCCCTTAAGGAGCCTGCTCCAACTACCTTAAGGAGCCTGCTCCAACTACCCCT 2079
QY 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82
Db 2080 AAGGAGCCTGCTCCAACTACCTTAAGGAGCTGCTCCAACTACCTTAAGGAGCCTGCTCCAACTACCCCT 2127
QY 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102
Db 2128 -----AAAGGAGCTGCTCCAACTACCTTAAGGAGCTGCTCCAACTACCTTAAGGAGCCTGCTCCAACTACCCCT 2175
QY 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
Db 2176 AAGAGCCTGCCCCCAAGGAGCTGTCACCCACCCACCAAGGAGCCTGTCACCCACCCACCCACCCACCC 2235
QY 114 GluGluGluAlaArgLeuLeuProArgThrHisGlnAlaGluLeuHisGlnHis 133
Db 2236 TGTGAAGAGCCCTCCAACTACCTTAAGGAGCTGCTCCAACTACCCCTAAGGAGCCT 2295
QY 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
Db 2296 GCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCTTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 2355
QY 154 ThrGlnGluValThrProLeuLeuGlnLysLeuProGluLeuValHisAla 173
Db 2356 CTCGAAGAACCTGCACCCCACTACTCCCAAGAGCCTGCCCCCAAGGAGCTTGACCCACC 2415
QY 174 ThrLeuSerThrPro 178
Db 2416 ACCACCAAGGGGCC 2430
RESULT 8
US-07-757-022B-43
; Sequence 43, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3813 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3810
US-07-757-022B-43
Alignment Scores:
Pred. No.: 0.0137 Length: 3813
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
Gaps: 4
US-10-030-225-2_COPY_27_213 (1-187) x US-07-757-022B-43 (1-3813)
```



Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1502 CCACCCCTGAGAACCCCGACCCACCCACCCCTGAG-GAGCTCGCACCCACCCACCCCTGAG 1560  
Qy 26 ProAspProArgProLeuLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 45  
Db 1561 ---GAGCCACACCCACCCACCCCTGAGGAGGCTGCTCCACCACTCCCAAGGAGCGGCT 1617  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1618 CCCAACACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1677  
Qy 65 -----ProGlyAenThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1678 AAGAGGCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT----- 1725  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1726 -----AAGGGAGCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1773  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1774 AAGAGGCTGCTCCCAAGGAGCTTGCACCCACCCACCCACCCACCCACCCACCCACCC 1833  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1834 TCTGACAAAGCCGCTCCAACTACCCCTAAGGGGAGCTGCTCCAACTACCCCTAAGGGGAGCT 1893  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAenAlaThrProProArg 153  
Db 1894 GCTCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGGGAGCTGCTCCAACTACCC 1953  
Qy 154 ThrGlnGluValThrProLeuLeuLeuLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1954 CTCAAGGAACCTGACCCACCTACTCCCAAGAGGCTGCTCCCAAGAGGCTTGCACCCACCC 2013  
Qy 174 ThrLeuSerThrPro 178  
Db 2014 ACCACCAAGGGGCC 2028

## RESULT 9

US-07-757-022B-41  
; Sequence 41, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/457,196  
;; FILING DATE: 29-DEC-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/390,901  
;; FILING DATE: 08-AUG-1989  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Cseri, Luann  
;; REGISTRATION NUMBER: 31,822  
;; REFERENCE/DOCKET NUMBER: GI 5190  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617)876-1170  
;; TELEFAX: (617)876-5851  
;; INFORMATION FOR SEQ ID NO: 41:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3936 base pairs  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: double  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..3933  
US-07-757-022B-41  
Alignment Scores:  
Pred. No.: 0.0143 Length: 3936  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4  
US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-41 (1-3936)  
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1625 CCACCCCTGAGAACCCCGACCCACCCACCCCTGAG-GAGCTCGCACCCACCCACCCCTGAG 1683  
Qy 26 ProAspProArgProLeuLysGluGluGluGluGluGluGluGluGluGluGluGluGlu 45  
Db 1684 ---GAGCCACACCCACCCACCCCTGAGGAGGCTGCTCCACCACTCCCAAGGAGCGGCT 1740  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1741 CCACACACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1800  
Qy 65 -----ProGlyAenThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1801 AAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT----- 1848  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1849 -----AAGGGAGCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT----- 1896  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1897 AAGAGGCTGCTCCCAAGGAGCTTGCACCCACCCACCCACCCACCCACCCACCCACCC 1956  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1957 TCTGACAAAGCCGCTCCAACTACCCCTAAGGGGAGCTGCTCCAACTACCCCTAAGGGGAGCT 2016  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAenAlaThrProProArg 153  
Db 2017 GCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGGGAGCTGCTCCAACTACCC 2076  
Qy 154 ThrGlnGluValThrProLeuLeuLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2077 CTCAAGGAACCTGACCCACCTACTCCCAAGAGGCTGCTCCCAAGAGGCTTGCACCCACCC 2136  
Qy 174 ThrLeuSerThrPro 178  
Db 174 ThrLeuSerThrPro 178

Tue Feb 22 09:49:23 2005

```
Db      2137 ACCACCAAGGGGCC 2151
RESULT 10
; Sequence 141, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07757,022B
; APPLICATION NUMBER: US 07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 141:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3942 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3939
; US-07-757-022B-141
Alignment Scores:
Pred. No.: 0.0143 Length: 3942
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 3 Gaps: 4
US-10-030-225-2_COPY_27_213 (1-187) x US-07-757-022B-141 (1-3942)
QY      6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db      1631 CCACCCCTGAGAGCCCGCCGACCACCCACCCCTGAG-GAGCTGCCACCCACCCCTGAG 1689
26 ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45
1690 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGAGGGCT 1746
46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
1747 CCCAACACCCCTAAGGAGGCTGCTCCCACTACCCCTAAGGAGGCTGCTCCAACTACCCCT 1806
65 -----ProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeuGlu 82
1807 AAGGAGCCTGCTCAACTACCCCTAAGGAGACTGCTCCAACTACCCCT----- 1854
83 LeuGlnLysLeuProGlyLeuAlaSerThrLeuSerThrProAsnProAspThrGln 102
1855 -----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCACTACTCCC 1902
103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
1903 AAGAAGCCTGCCCCCAAGGAGCTTGACCCACCACCAAGAGGCCACATCCACCACC 1962
114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluHisGlnHis 133
1963 TCTGACAAAGCCCGCTCAACTACCCCTAAGGGGACTGCTCCAACTACCCCTAAGGAGCCT 2022
134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
2023 GCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGGGACTGCTCCAACTACC 2082
154 ThrGlnGluValThrProLeuLeuGluGlnLysLeuProGluLeuValHisAla 173
2083 CTCAAGGAACCTGCCACCCACTACTCTCCCAAGAGCCTGCCCCCAAGGAGCTTGACCCACC 2142
174 ThrLeuSerThrPro 178
2143 ACCACCAAGGGGCC 2157
RESULT 11
US-07-757-022B-49
; Sequence 49, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 141:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3942 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3939
; US-07-757-022B-141
Alignment Scores:
Pred. No.: 0.0143 Length: 3942
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 3 Gaps: 4
US-10-030-225-2_COPY_27_213 (1-187) x US-07-757-022B-141 (1-3942)
QY      6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db      1631 CCACCCCTGAGAGCCCGCCGACCACCCACCCCTGAG-GAGCTGCCACCCACCCCTGAG 1689
```

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 49:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3945 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3942  
US-07-757-022B-49

Alignment Scores:  
Pred. No.: 0.0143 Length: 3945  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-49 (1-3945)

Qy	6	ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer	25
Db	1634	CCACCCCTGAGAGCCCGCCACCACCCCTGAG-GAGCTCGCACCCACCCCTGAG	1692
Qy	26	ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu	45
Db	1693	--GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGCGGCT	1749
Qy	46	GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMet	64
Db	1750	CCCAACACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT	1809
Qy	65	-----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu	82
Db	1810	AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT	1857
Qy	83	LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln	102
Db	1858	-----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCCACTACTCC	1905
Qy	103	AlaSerAlaSerPro-----AspProArgProLeuArg	113
Db	1906	AAGAGCTGCTGCCCCAAGGAGCTTGCACCCACCCACCAAGGAGCCACCACTCCACCCACC	1965
Qy	114	GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis	133
Db	1966	TCTGACAGCCGCTCCAACTACCCCTAAGGGAGCTGCTCAACTACCCCTAAGGAGCT	2025
Qy	134	GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg	153
Db	2026	GCTCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACC	2085
Qy	154	ThrGlnGluValThrProLeuLeuGlnLysLeuProGluLeuValHisAla	173
Db	2086	CTCAAGGAACCTGCACCCCACTACTCCCAAGAGCCTGCTCCCAAGAGCCTTGCACCCACC	2145
Qy	174	ThrLeuSerThrPro	178
Db	2146	ACCACCAAGGGGCC	2160

RESULT 12

US-07-757-022B-45  
; Sequence 45, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3963 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3960  
US-07-757-022B-45

Alignment Scores:  
Pred. No.: 0.0144 Length: 3963  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-45 (1-3963)

Qy	6	ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer	25
Db	1652	CCACCCCTGAGAGCCCGCCACCACCCCTGAG-GAGCTCGCACCCACCCCTGAG	1710
Qy	26	ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu	45

Db 1711 ---GAGCCACACCCACCCAGGAGGCTGCTCCACCACTCCCAAGGACGCGCT 1767  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1768 CCCAACACCCCTAAGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1827  
Qy 65 -----ProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1828 AAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT----- 1875  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1876 -----AAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1923  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1924 AAGAGCGCTGCTCCCAAGGAGCTTCCACCCACCCACCAAGGAGCGCCACATCCACCAACC 1983  
Qy 114 GluGluGluGluAlaArgLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1984 TCTGACACGCGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCT 2043  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 2044 GCTCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACC 2103  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2104 CTCAAGGAACCTGCACCCCACTACTCCCAAGAGCGCTGCTCCCAAGGAGCGCTGCAACCCACC 2163  
Qy 174 ThrLeuSerThrPro 178  
Db 2164 ACCACCAAGGGGCGCC 2178

## RESULT 13

US-07-757-022B-59  
; Sequence 59, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3963 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3960  
US-07-757-022B-59  
Alignment Scores:  
Pred. No.: 0.0144 Length: 3963  
Scores: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 3 Gaps: 4  
US-10-030-225-2\_COPY\_27\_213 (1-187) x US-07-757-022B-59 (1-3963)  
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1652 CCACCCCTGAGAGCGCGCACCCACCCCTGAG-GAGCTCGCACCCACCCCTCGAG 1710  
Qy 26 ProAspProArgProLeuLysGluGluGluGluProLeuLeuProArgThrHisLeu 45  
Db 1711 ---GAGCCACACCCACCCACCCCTGAGGAGCGCTGCTCCCACTCCCAAGGAGCGCGCT 1767  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1768 CCCAACACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1827  
Qy 65 -----ProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1828 AAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCCACTACCCCT----- 1875  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1876 -----AAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCAACCCACTACTCCC 1923  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1924 AAGAGCGCTGCTCCCAAGGAGCTTCCACCCACCCACCAAGGAGCGCCACATCCACCAACC 1983  
Qy 114 GluGluGluGluAlaArgLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1984 TCTGACACGCGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCT 2043  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 2044 GCTCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACC 2103  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2104 CTCAAGGAACCTGCACCCCACTACTCCCAAGAGCGCTGCTCCCAAGGAGCGCTTGCACCCACC 2163  
Qy 174 ThrLeuSerThrPro 178  
Db 2164 ACCACCAAGGGGCGCC 2178  
RESULT 14  
US-07-757-022B-47  
; Sequence 47, Application US/07757022B  
; Patent No. 6433142

```

; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4065 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4062
; US-07-757-022B-47
;
; Alignment Scores:
; Pred. No.: 0.0149 Length: 4065
; Score: 144.00 Matches: 53
; Percent Similarity: 35.14% Conservative: 12
; Best Local Similarity: 28.65% Mismatches: 99
; Query Match: 14.47% Indels: 22
; DB: 3 Gaps: 4
;
; US-10-030-225-2_COPY_27_213 (1-187) x US-07-757-022B-47 (1-4065)
;
; Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
; Db 1754 CCACCCCTGAGAGCCCGCACCCACCCCTGAG-GAGCTCGCACCCACCCACCCCTGAG 1812
; Qy 26 ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45
; Db 1813 ---GAGCCACACCCACCCACCCCTGAGGAGGCTGCTCCACCATCCCAAGGAGCGGCT 1869
; Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
;
; Db 1870 CCCAACACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 1929
; Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82
; Db 1930 AAGGAGCGCTGCTCCAACTACCCCTAAGGAGACTGCTCCAACTACCCCT----- 1977
; Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102
; Db 1978 -----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGACCCACTACTCCC 2025
; Qy 103 AlaSerAlaSerPro-----AppProArgProLeuArg 113
; Db 2026 AAGAAGCGTCCCGCCCAAGAGCTTGACCCACCCACCAAGGAGCCACATCCACCACC 2085
; Qy 114 GluGluGluAlaAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
; Db 2086 TCTGACAAGCCGCTCCAACTACCCCTAAGGGGACTGCTCCAACTACCCCTAAGGAGCGCT 2145
; Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
; Db 2146 GCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGGGACTGCTCCAACTACC 2205
; Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173
; Db 2206 CTCAAGGAACCTGCACCCACTACTCCCAAGGAAGCGCTGCCCAAGGAGCTTGACCCACC 2265
; Qy 174 ThrLeuSerThrPro 178
; Db 2266 ACCACCAAGGGGCGCC 2280
;
; RESULT 15
; US-07-757-022B-39
; Sequence 39, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4065 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..4062
; US-07-757-022B-47
;
; Alignment Scores:
; Pred. No.: 0.0149 Length: 4065
; Score: 144.00 Matches: 53
; Percent Similarity: 35.14% Conservative: 12
; Best Local Similarity: 28.65% Mismatches: 99
; Query Match: 14.47% Indels: 22
; DB: 3 Gaps: 4
;
; US-10-030-225-2_COPY_27_213 (1-187) x US-07-757-022B-47 (1-4065)
;
; Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
; Db 1754 CCACCCCTGAGAGCCCGCACCCACCCCTGAG-GAGCTCGCACCCACCCACCCCTGAG 1812
; Qy 26 ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45
; Db 1813 ---GAGCCACACCCACCCACCCCTGAGGAGGCTGCTCCACCATCCCAAGGAGCGGCT 1869
; Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64

```

us-10-030-225-2 copy 27 213.rni

Search completed: February 22, 2005, 03:20:11  
Job time : 464 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 02:21:12 ; Search time 5476 Seconds  
(without alignments)  
201.838 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995  
Sequence: 1 LPVKKPLRGRPGSLRRLA.....PELVHATLSTPNPDNQVTK 187

Scoring table:  
BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5384158 seqs, 2955248155 residues

Total number of hits satisfying chosen parameters: 10768316

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10030225/runat\_18022005\_100357\_2456/app.query.fasta\_1.327  
-DB=published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10030225 @CEN 1 1 723 @runat\_18022005\_100357\_2456  
-NCPU=6 -ICPU=3 -NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THRAADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No' is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	658.5	66.2	497	9	US-09-864-761-156 Sequence 156, App
c 2	430	43.2	243	9	US-09-864-761-16984 Sequence 16984, A
c 3	149.5	15.0	6694	17	US-10-221-613-274 Sequence 274, App
	149	15.0	1602	17	US-10-425-114-14755 Sequence 14755, A
c 5	147	15.0	1782	17	US-10-424-599-22752 Sequence 22752, A
c 6	147	14.8	6456	15	US-10-311-455-979 Sequence 979, App
c 7	144	14.5	2824	13	US-10-124-557-13 Sequence 13, Appl
c 8	144	14.5	3066	13	US-10-124-557-83 Sequence 83, Appl
c 9	144	14.5	3117	13	US-10-124-557-73 Sequence 73, Appl
c 10	144	14.5	3148	13	US-10-124-557-57 Sequence 57, Appl
c 11	144	14.5	3420	13	US-10-124-557-103 Sequence 103, App
c 12	144	14.5	3813	13	US-10-124-557-43 Sequence 43, Appl
c 13	144	14.5	3936	13	US-10-124-557-41 Sequence 41, Appl
c 14	144	14.5	3942	13	US-10-124-557-141 Sequence 141, App
c 15	144	14.5	3945	13	US-10-124-557-49 Sequence 49, Appl
c 16	144	14.5	3963	13	US-10-124-557-45 Sequence 59, Appl
c 17	144	14.5	3963	13	US-10-124-557-59 Sequence 47, Appl
c 18	144	14.5	4065	13	US-10-124-557-47 Sequence 39, Appl
c 19	144	14.5	4086	13	US-10-124-557-39 Sequence 51, Appl
c 20	144	14.5	4092	13	US-10-124-557-51 Sequence 61, Appl
c 21	144	14.5	4215	13	US-10-124-557-61 Sequence 303, App
c 22	144	14.5	4575	13	US-10-044-090-303 Sequence 1, Appli
c 23	144	14.5	5008	13	US-10-124-557-1 Sequence 1, Appli
c 24	144	14.5	5041	10	US-09-918-624B-1 Sequence 2, Appli
c 25	144	14.5	5041	11	US-09-897-188-2 Sequence 981, App
c 26	144	14.5	5041	17	US-10-641-643-981 Sequence 2, Appli
c 27	144	14.5	5041	18	US-10-858-595-2 Sequence 14, Appl
c 28	144	14.5	5041	18	US-10-473-974-14 Sequence 89, Appl
c 29	144	14.5	5041	18	US-10-473-974-89 Sequence 90, Appl
c 30	144	14.5	5041	18	US-10-473-974-90 Sequence 931, App
c 31	143	14.4	14643	15	US-10-156-761-931 Sequence 2, Appli
c 32	143	14.4	31422	18	US-10-204-862A-2 Sequence 1, Appli
c 33	143	14.4	9025608	15	US-10-156-761-1 Sequence 1, Appli
c 34	142.5	14.3	88421	9	US-09-976-059-1 Sequence 94182, A
c 35	136.5	13.7	1257	18	US-10-437-963-94182 Sequence 7594, Ap
c 36	136.5	13.7	11294	10	US-09-764-891-7594 Sequence 7595, Ap
c 37	136.5	13.7	15123	10	US-09-764-891-7595 Sequence 382, App
c 38	136.5	13.7	75853	13	US-10-087-192-382 Sequence 88274, A
c 39	135	13.6	2149	18	US-10-437-963-88274 Sequence 700, App
c 40	135	13.6	34722	18	US-10-322-281-700 Sequence 1421, Ap
c 41	134.5	13.5	2626	17	US-10-108-260A-1421 Sequence 36853, A
c 42	133	13.4	885	18	US-10-363-345A-36853 Sequence 36854, A
c 43	133	13.4	885	18	US-10-363-345A-36854 Sequence 3, Appli
c 44	133	13.4	92638	18	US-10-450-826-3 Sequence 47999, A
c 45	132.5	13.3	3725	18	US-10-437-963-47999

ALIGNMENTS

RESULT 1

US-09-864-761-156/c  
; Sequence 156, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Acomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864, 761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03

```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 156
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF111168.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.85
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.75
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.97
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; US-09-864-761-156

Alignment Scores:
Pred. No.: 5,248-55 Length: 497
Score: 658.50 Matches: 126
Percent Similarity: 76.97% Conservative: 1
Best Local Similarity: 76.36% Mismatches: 37
Query Match: 66.18% Indels: 37
DB: 9 Gaps: 1

US-10-030-225-2_COPY_27_213 (1-187) x US-09-864-761-156 (1-497)

QY 32 LysGluGluGluAlaProLeuProArgThrHisLeuGlnAlaGluProHisGln 51
Db 496 AGGAAGAGGAGGAGGACCACCTGCTCCCGAAGCCACCTGCAGGAGGACACACCAA 437
QY 52 HisGlyCysTrpThrValThrGluProAlaAlaMetThrProGlyAsnThrThrProPro 71
Db 436 CATGATGTGCTGCTCACTGAGCAGCAGCATGACCCAGGCAACCCACCTCCCTCCC 377
QY 72 ArgThrProGluValThrProLeuArgLeuGluGlnLysLeuProGlyLeuAlaSer 91
Db 376 AGGACCCAGAGTTACTCGTGGCTGGAGCTGCAGAAAGCTGCCGGGATGGCCAAAC 317
QY 92 ThrThrLeuSerThrProAsnProAspThr----- 101
```

```
Db 316 ACAACCTTGAGTACCCCTTACCTGATACCCAGGTGAGAGCTACAGAAGGCCAGCAGCT 257
QY 101 ----- 101
Db 256 CTGGGCTCTGCCTGGCCAGCCTTTTCTGGGTCCCTGTGGTAAAGCCCTAACCTGTCACC 197
QY 102 -----GlnAlaSerAlaSerProAspProArgProLeuArgGlu 114
Db 196 TCACCTTCCCTCTCTCTCTCCTCAGGCTTCAAGCTCCCCAGATCCTAGGCCTCTGAGGAA 137
QY 115 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGly 134
Db 136 GAGGAGGAGGAGCAGACTGCTCCCGAAGCCACCTGCAGGAGCTACACCAATGGA 77
QY 135 CysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArgThr 154
Db 76 TGTGGACTGTCACTGAGCAGCAGGCTGACCCAGGGAATGCCAGCCTCCAGGACC 17
QY 155 GlnGluValThrPro 159
Db 16 CAGGAGGTACTCCC 2

RESULT 2
US-09-864-761-16984/c
; Sequence 16984, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
```



[illegible]

```

; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 14755
; LENGTH: 1602
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LfB3039-011-B5_FLI
US-10-425-114-14755

```

Alignment Scores:	
Pred. No.:	5.57e-05
Score:	149.00
Percent Similarity:	40.00%
Best Local Similarity:	30.48%
Query Match:	14.97%
DB:	17
	Gaps:
	8
	Indels:
	48
	Mismatches:
	78
	Conservative:
	20
	Matches:
	64
	Length:
	1602

US-10-030-225-2 COPY 27 213 (1-187) x US-10-425-114-14755 (1-1602)

11	ProArgProGlySerLeuThrArgLeuAlaGlu---ValSerAlaSerProAspProArg	29
111	CCAGCTCCCTCTCTCCAAACGCCGCGCATCTCCCTCCCTCCCTCCGCTG	170
30	ProLeuLysGluGluGlu-----GluAlaPro	38
171	CCCTCTCTCTTTTCCACATCGCTCCGCTCTCTCTCGTCGCCTCCACGGCGCCT	230
39	LeuLeuProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThr	58
231	CTCTCTCTCGGACT-----ACTCTGTCT	254
59	GluProAlaAlaMetThrProGlyAsnThrThrProPro-----ArgThr	73
255	CTGCCTTCGCGACCTCTCCACAGAACACATCTGGTCCCTGGAGCCCTCTCCGCGAGC	314
74	ProGluValThrProLeuArgLeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThr	93
315	CCCACTGGCGCTCCCTCTACGGCTCTCTTACCCCGCTTCACCACTGTGTGACAAAGC	374
94	LeuSer-----ThrProAsnProAspThrGlnAla-----	103
375	TCAAGCCCCACATCGCCCTCTCCAACCTCTCCCTCCCTCCGACTACGCGTGCCTG	434
104	SerAla-SerProAspProArgProLeuArgGluGluGluAlaArgLeuLeuLeuProAr	123
435	TCTCTCCGCGCTCGCCACAGCGCTCTCCGCCAAACCCCTCGCCTCCCGTACTCTCCCTG	494
123	gThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGluProAla	143
495	ACCCCTACCTCGTCTCCAAGATCACCAACATGGTCAACCCGCTCTCGCCACCAAGCTCT	554
143	a-----LeuThrProGlyAsnAlaThrProPro-ArgThrGlnGluValThrProLeu	161
555	ACCTGAGTTCATCAAGATCCCGCTCGCGCCGCCGCCCTCTCTCGAAACCAACCAAGCCT	614
161	euLeuLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrPro-----	178
615	TCGAGGAAGTCACTCTCGTCTCCCAACATGTGGCGGCCATCGACACCAACCCCGTCCACC	674
179	-----AsnProAspAsnGlnVal	184
675	TCCGATTAAACCCCTAACCCCTAACCTT	700

## RESULT 5

US-10-424-599-22752  
; Sequence 22752, Application US/10424599  
; Publication No. US20040031072A1

```

; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 22752
; LENGTH: 1782
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_120547C.1
; US-10-424-599-22752

```

Alignment Scores:	
Pred. No.:	6.25e-05
Score:	149.00
Percent Similarity:	40.00%
Best Local Similarity:	30.48%
Query Match:	14.97%
DB:	17
	17
	8
Length:	1782
Matches:	64
Conservative:	20
Mismatches:	78
Indels:	48
Gaps:	8

US-10-030-225-2 COPY 27 213 (1-187) x US-10-424-599-22752 (1-1782)

Qy	11	ProArgProGlySerLeuThrArgLeuAlaGlu---ValSerAlaSerProAspProArg	29
Db	138		197
Qy	30	ProLeuLysGluGluGlu-----GluAlaPro	38
Db	198		257
Qy	39	LeuLeuProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThr	58
Db	258		281
Qy	59	GluProAlaAlaMetThrProGlyAsnThrThrProPro-----ArgThr	73
Db	282		341
Qy	74	ProGluValThrProLeuArgLeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThr	93
Db	342		401
Qy	94	LeuSer-----ThrProAsnProAspThrGlnAla-----	103
Db	402		461
Qy	104	SerAla-SerProAspProArgProLeuArgGluGluGluAlaArgLeuLeuProAr	123
Db	462		521
Qy	123	gThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGluProAla	143
Db	522		581
Qy	143	a-----LeuThrProGlyAsnAlaThrProPro-ArgThrGlnGluValThrProLeu	161
Db	582		641
Qy	161	eUeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrPro-----	178
Db	642		701
Qy	179	-----AsnProAspAsnGlnVal	184
Db	702		727



Db 1307 CCACCTGAGAGCCCGACCCACCCCTGAG-GAGCTGCGACCCACCCCTGTGAG 1365  
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuProArgThrHisLeu 45  
Db 1366 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACCACTCCCAAGGAGCGCT 1422  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1423 CCAACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1482  
Qy 65 -----ProGlyAsnThrProProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1483 AAGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1530  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1531 -----AAAGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCCACTACTCCC 1578  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1579 AAGAGCTGCTCCCAAGGAGCTGACCCACCCACCCCAAGGAGCCCACTCCACACACC 1638  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1639 TCTGACAGCCGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCT 1698  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 1699 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACC 1758  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 1759 CTCAGGAACCTGCACCACTACTCCCAAGAGCCCTGCCCCCAAGGAGCTTGCACCCACC 1818  
Qy 174 ThrLeuSerThrPro 178  
Db 1819 ACCACCAAGGGGCC 1833  
RESULT 8  
US-10-124-557-83  
; Sequence 83, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/10/124,557  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 83:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3066 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3066  
; SEQUENCE DESCRIPTION: SEQ ID NO: 83:  
US-10-124-557-83  
Alignment Scores:  
Pred. No.: 0.000344 Length: 3066  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
Gaps: 4  
US-10-030-225-2\_COPY\_27\_213 (1-187) x US-10-124-557-83 (1-3066)  
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1550 CCACCTGAGAGCCCGACCCACCCCTGAG-GAGCTGCGACCCACCCCTGTGAG 1608  
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuProArgThrHisLeu 45  
Db 1609 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCCACTACCCCTCAAGGAGCGCT 1665  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1666 CCCAACACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT 1725  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1726 AAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1773  
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1774 -----AAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTACTCCC 1821  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1822 AAGAGCTGCTGCCCCCAAGGAGCTTGCACCCACCCACCCCAAGGAGCCCACTCCACACC 1881  
Qy 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1882 TCTGACAGCCGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCT 1941  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153  
Db 1942 GCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACC 2001  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2002 CTCAGGAACCTGCACCACTACTCCCAAGAGCTGCCCCCAAGGAGCTTGCACCCACC 2061  
Qy 174 ThrLeuSerThrPro 178  
Db 2062 ACCACCAAGGGGCC 2076

	Db	1561	--GAGCCACACCCACCGCTTGAAGAGCCTGTCTCCCAACTCCCAAGGCAGGGCT	161
	Qy	46	GlnAlaGluProHisGlnHisGlyCysThrValThrGluProAlaAlaMetThr---	64
	Db	1618	CCCACACCCTTAAGAGCCTGTCTCCAACCTAACCCCTAAGGAGCCTGCTCCAATCACCT	1677
	Qy	65	-----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu	82
	Db	1678	AAGGAGCCTGTCTCCAACCTAGCCCTTAAGGAGACTGCTCCAATCACCT	1725
	Qy	83	LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln	102
	Db	1726	-----AAAGGAGCTGTCTCCAACCTCAAGGAACCTGCACCCACTACTCCTCC	1773
	Qy	103	AlaSerAlaSerPro-----AspProArgProLeuArg	113
	Db	1774	AAGAAGCCTGCCCCCAAGGAGCTTGACACCACCAAGGAGCCCATCCACCACC	1833
	Qy	114	GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis	133
	Db	1834	TCTGACAAGCCGCTCTCCAACCTAGCCCTTAAGGAGACTGCTCCAATCACCTAAGGAGCCT	1893
	Qy	134	GlyCysThrThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg	153
	Db	1894	GCTCAACTACCCCTTAAGGAGCCTGTCTCCAACCTAACCCCTAAGGGGACTGCTCCAATCACCT	1953
	Qy	154	ThrGlnGluValThrProLeuLeuGlnLysLeuProGluLeuValHisAla	173
	Db	1954	CTCAAGGAACTGCACCCACTACTCCCAAGAGCCTGCCCAAGAGGCTTGACACCCACC	2013
	Qy	174	ThrLeuSerThrPro 178	
	Db	2014	ACCACCAAGGGGCC 2028	
		RESULT 10		
		US-10-124-557-57		
		; Sequence 57, Application US/10124557		
		; Publication No. US20020137894A1		
		GENERAL INFORMATION:		
		APPLICANT: Turner, Katherine		
		Clark, Stephen C.		
		Jacobs, Kenneth		
		Hewick, Rodney M.		
		Gesner, Thomas G.		
		TITLE OF INVENTION: Megakaryocyte Stimulating Factors		
		NUMBER OF SEQUENCES: 143		
		CORRESPONDENCE ADDRESS:		
		ADDRESSEE: Genetics Institute, Inc.		
		STREET: 87 CambridgePark Drive		
		CITY: Cambridge		
		STATE: Massachusetts		
		COUNTRY: U.S.A.		
		ZIP: 02140		
		COMPUTER READABLE FORM:		
		MEDIUM TYPE: Floppy disk		
		COMPUTER: IBM PC compatible		
		OPERATING SYSTEM: PC-DOS/MS-DOS		
		SOFTWARE: PatentIn Release #1.0, Version #1.25		
		CURRENT APPLICATION DATA:		
		APPLICATION NUMBER: US/10/124,557		
		FILING DATE: 16-Apr-2002		
		CLASSIFICATION: <Unknown>		
		PRIOR APPLICATION DATA:		
		APPLICATION NUMBER: US 07/643,502		
		FILING DATE: 18-JAN-1991		
		APPLICATION NUMBER: US 07/546,114		
		FILING DATE: 29-JUN-1990		
		APPLICATION NUMBER: US 07/457,196		
		FILING DATE: 29-DEC-1989		
		APPLICATION NUMBER: US 07/390,901		
		FILING DATE: 08-AUG-1989		
		ATTORNEY/AGENT INFORMATION:		
		NAME: Cserr, Luann		
		REGISTRATION NUMBER: 31,822		
		REFERENCE/DOCKET NUMBER: GI 5190		
		TELECOMMUNICATION INFORMATION:		
		TELEPHONE: (617)876-1170		
		TELEFAX: (617)876-5851		
		INFORMATION FOR SEQ ID NO: 73:		
		SEQUENCE CHARACTERISTICS:		
		LENGTH: 3117 base pairs		
		TYPE: nucleic acid		
		STRANDEDNESS: double		
		TOPOLOGY: unknown		
		MOLECULE TYPE: cDNA		
		FEATURE:		
		NAME/KEY: CDS		
		LOCATION: 1..3114		
		SEQUENCE DESCRIPTION: SEQ ID NO: 73:		
		US-10-124-557-73		
		Alignment Scores:		
		Pred. No.: 0.00035 Length: 3117		
		Score: 144.00 Matches: 53		
		Percent Similarity: 35.14% Conservativeness: 12		
		Best Local Similarity: 28.65% Mismatches: 99		
		Query Match: 14.47% Indels: 22		
		DB: 13 Gaps: 4		
		US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-73 (1-3117)		
	Qy	6	ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25	
	Db	1502	CCACCCCTGAGAGCCCGCCACCCACCCCTGAG-GAGTCGACCCACCCACCCCTGAG 1560	
	Qy	26	ProArgProArgProLeuLysGluGluAlaProLeuArgThrHisLeu 45	

```
;
; REGISTRATION NUMBER: 31,922
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3148 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3147
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-10-124-557-57

Alignment Scores:
Pred. No.: 0.000354 Length: 3148
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4

US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-57 (1-3148)
QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1631 CCACCCCTGAGAGCGCCGACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1689
QY 26 ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1690 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACACTCCCAAGGCGAGCGCT 1746
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr 64
Db 1747 CCCACACCCCTAAGGAGCGCTGCTCCACTACCCCTAAGGAGCGCTGCTCCACTACCCCT 1806
QY 65 -----ProGlyAsnThrThrProProArgThrProGluValThrProLeuArgLeuGlu 82
Db 1807 AAGGAGCGCTGCTCCAACTACCCCTAAGGAGACTGCTCCAACTACCCCT----- 1854
QY 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102
Db 1855 -----AAGGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCGACCACTACCTCCC 1902
QY 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
Db 1903 AAGAGCGCTGCCCCCAAGGAGCTTGCACCCACCCACCAAGGAGCCCATCCACCAACC 1962
QY 114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db 1963 TCTGCAAGCGCGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACCCCTAAGGAGCGCT 2022
QY 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153
Db 2023 GCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACC 2082
QY 154 ThrGlnGluValThrProLeuLeuLeuGluGlnLysLeuProGluLeuValHisAla 173
Db 2083 CTCAGGAACCTGCACCCACTACTCCCAAGAAGCGCTGCCCCCAAGGAGCTTGCACCCACC 2142
QY 174 ThrLeuSerThrPro 178
Db 2143 ACCACCAAGGGGCC 2157

RESULT 11
US-10-124-557-103
; Sequence 103, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseriz, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3420 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3420
; SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-10-124-557-103

Alignment Scores:
Pred. No.: 0.000387 Length: 3420
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4

US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-103 (1-3420)
QY 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1904 CCACCCCTGAGAGCGCCGACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1962
QY 26 ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1963 ---GAGCCACACCCACCCCTGAGGAGCGCTGCTCCACCACTCCCAAGGCGAGCGCT 2019
QY 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr 64
Db 2020 CCCAACACCCCTAAGGAGCGCTGCTCCAACTACCCCTAAGGAGCGCTGCTCCAACTACCCCT 2079
```

```
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82
Db 2080 AAGGAGCCTGCTCAACTACCCCTAAGGAGACTGCTCCAACTACCCCT----- 2127
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrLeuSerThrProAsnProAspThrGln 102
Db 2128 -----AAGGAGCTGCTCAACTACCCCTAAGGAGAACTGACCCACTACTCCC 2175
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
Db 2176 AAGAGCCTGCCCCCAAGGAGCTTGACCCACCACCCAGGAGCCACATCCACCACC 2235
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db 2236 TCTGACAAAGCCGCTCCAACTACCCCTAAGGGGAGCTGCTCCAACTACCCCTAAGGAGCCT 2295
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
Db 2296 GCTCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGGGAGCTGCTCCAACTACC 2355
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173
Db 2356 CTCAAGGAAGCTGACCCACTACTCCCAAGAGCCTGCCCCCAAGGAGCTTGACCCACC 2415
Qy 174 ThrLeuSerThrPro 178
Db 2416 ACCACCAAGGGGCC 2430

RESULT 12
US-10-124-557-43
; Sequence 43, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ceert, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
```

```
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3813 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3810
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-10-124-557-43
```

```
Alignment Scores:
Pred. No.: 0.000436 Length: 3813
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4
```

US-10-030-225-2\_copy\_27\_213 (1-187) x US-10-124-557-43 (1-3813)

```
Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1502 CCACCCTGAGAGCCGCGCACCCACCACCCCTGAG-GAGCTCGCACCCACCACCCCTGAG 1560
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProLeuArgThrHisLeu 45
Db 1561 ---GAGCCCAACACCCACCACCCCTGAGGAGCTGCTCCACCACCTCCCAAGGAGCGCT 1617
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
Db 1618 CCCAACACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1677
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82
Db 1678 AAGGAGCCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCTAAGGAGCCTGCTCC 1725
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102
Db 1726 -----AAAGGAGCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1773
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
Db 1774 AAGAGCCTGCCCCCAAGGAGCTTGACCCACCACCACCCACATCCACACCACC 1833
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db 1834 TCTGACAAAGCCGCTCCAACTACCCCTAAGGGGAGCTGCTCCAACTACCCCTAAGGAGCCT 1893
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
Db 1894 GCTCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGGGAGCTGCTCCAACTACC 1953
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173
Db 1954 CTCAGGAACCTGCACCCACTACTCCCAAGAGCCTGCCCCCAAGGAGCTTGACCCACCACC 2013
Qy 174 ThrLeuSerThrPro 178
Db 2014 ACCACCAAGGGGCC 2028
```

## RESULT 13

US-10-124-557-41  
; Sequence 41, Application US/10124557  
; Publication No. US20020137894A1

## GENERAL INFORMATION:

APPLICANT: Turner, Katherine  
Clark, Stephen C.  
Jacobs, Kenneth  
Hewick, Rodney M.  
Gesner, Thomas G.

;; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
;; NUMBER OF SEQUENCES: 143  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genetics Institute, Inc.  
;; STREET: 87 CambridgePark Drive  
;; CITY: Cambridge  
;; STATE: Massachusetts  
;; COUNTRY: U.S.A.  
;; ZIP: 02140

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/124,557  
;; FILING DATE: 16-Apr-2002  
;; CLASSIFICATION: <Unknown>  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/643,502  
;; FILING DATE: 18-JAN-1991  
;; APPLICATION NUMBER: US 07/546,114  
;; FILING DATE: 29-JUN-1990  
;; APPLICATION NUMBER: US 07/457,196  
;; FILING DATE: 29-DEC-1989  
;; APPLICATION NUMBER: US 07/390,901  
;; FILING DATE: 08-AUG-1989

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Geert, Luann  
;; REGISTRATION NUMBER: 31,822  
;; REFERENCE/DOCKET NUMBER: GI 5190  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617)876-1170  
;; TELEFAX: (617)876-5851

;; INFORMATION FOR SEQ ID NO: 41:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3936 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..3933  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 41:

US-10-124-557-41

Alignment Scores:  
Pred. No.: 0.000451 Length: 3936  
Score: 144.00 Matches: 53  
Percent Similarity: 35.14% Conservative: 12  
Best Local Similarity: 28.65% Mismatches: 99  
Query Match: 14.47% Indels: 22  
DB: 13 Gaps: 4

US-10-030-225-2\_COPY\_27\_213 (1-187) x US-10-124-557-41 (1-3936)

Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25  
Db 1625 CCACCCCTGAGAAGCCGACCCACCCCTGAG-GAGCTGGCACCACCCACCCCTGAG 1683  
Qy 26 ProAspProArgProLeuLysGluGluAlaProLeuLeuProArgThrHisLeu 45  
Db 1684 ---GAGCCACACCCACCCCTGAGGAGCTGTCTCCACCACTCCCAAGGACGGCT 1740  
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64  
Db 1741 CCCAACACCCCTAAGGAGCTGTCTCCAACTACCCCTAAGGAGCTGTCTCCAACTACCCCT 1800  
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82  
Db 1801 AAGGAGCTGTCTCCAACTACCCCTAAGGAGCTGTCTCCAACTACCCCT----- 1848

Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102  
Db 1849 -----AAAGGAGCTGTCTCCAACTACCCCTAAGGAGCTGTCTCCAACTACCCCT 1896  
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113  
Db 1897 AAGAAAGCTGCCCCCAAGGAGCTTGCACCCACCAAGGAGCTGTCTCCAACTACCCCT 1956  
Qy 114 GluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133  
Db 1957 TCTGACAGCCGCTCCAACTACCCCTAAGGAGCTGTCTCCAACTACCCCTAAGGAGCT 2016  
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153  
Db 2017 GCTCAACTACCCCTAAGGAGCTGTCTCCAACTACCCCTAAGGAGCTGTCTCCAACTAC 2076  
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173  
Db 2077 CTCAGGAACCTGCCACCCACTCTCCCAAGAAAGCTGTCTCCCAAGGAGCTGTCTCCAC 2136  
Qy 174 ThrLeuSerThrPro 178  
Db 2137 ACCACCAAGGGGCC 2151

#### RESULT 14

US-10-124-557-141  
; Sequence 141, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:

;; APPLICANT: Turner, Katherine  
;; Clark, Stephen C.  
;; Jacobs, Kenneth  
;; Hewick, Rodney M.  
;; Gesner, Thomas G.

;; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

;; NUMBER OF SEQUENCES: 143

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Genetics Institute, Inc.

;; STREET: 87 CambridgePark Drive

;; CITY: Cambridge

;; STATE: Massachusetts

;; COUNTRY: U.S.A.

;; ZIP: 02140

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patent In Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/10/124,557

;; FILING DATE: 16-Apr-2002

;; CLASSIFICATION: <Unknown>

;; PRIORITY APPLICATION DATA:

;; APPLICATION NUMBER: US 07/643,502

;; FILING DATE: 18-JAN-1991

;; APPLICATION NUMBER: US 07/546,114

;; FILING DATE: 29-JUN-1990

;; APPLICATION NUMBER: US 07/457,196

;; FILING DATE: 29-DEC-1989

;; APPLICATION NUMBER: US 07/390,901

;; FILING DATE: 08-AUG-1989

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Geert, Luann

;; REGISTRATION NUMBER: 31,822

;; REFERENCE/DOCKET NUMBER: GI 5190

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (617)876-1170

;; TELEFAX: (617)876-5851

;; INFORMATION FOR SEQ ID NO: 141:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 3942 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: double



```
; ; TOPOLOGY: unknown
; ; MOLECULE TYPE: cDNA
; ; FEATURE:
; ; NAME/KEY: CDS
; ; LOCATION: 1..3939
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 141:
US-10-124-557-141

Alignment Scores:
Pred. No.: 0.000452 Length: 3942
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4

US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-141 (1-3942)

Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1631 CCACCCCTGAGAAGCCCGCACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1689
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1690 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACACTCCCAAGCAGCGGCT 1746
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
Db 1747 CCCAACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1806
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82
Db 1807 AAGGAGCCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1854
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102
Db 1855 -----AAGGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCACTACTCCC 1902
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
Db 1903 AAGAAGCCTGCCCCCAAGGAGCTTGACCCACCCACCCACCCACCCACCATCCACCACC 1962
Qy 114 GluGluGluAlaArgLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db 1963 TCTGACAGCCGCTGCCAACTACCCCTAAGGGAGCTGCTCCAACTACCCCTAAGGAGCCT 2022
Qy 134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProArg 153
Db 2023 GCTCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGGAGCTGCTCCAACTACC 2082
Qy 154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173
Db 2083 CTCAGGAACCTGCACCCACTACTCCCAAGAGCCTGCCCCCAAGGAGCTTGACCCACC 2142
Qy 174 ThrLeuSerThrPro 178
Db 2143 ACCACCAAGGGGCC 2157

RESULT 15
US-10-124-557-49
; Sequence 49: Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; ; Clark, Stephen C.
; ; Jacobs, Kenneth
; ; Hewick, Rodney M.
; ; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Bridgepark Drive
```

```
; ; CITY: Cambridge
; ; STATE: Massachusetts
; ; COUNTRY: U.S.A.
; ; ZIP: 02140
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: PatentIn Release #1.0, Version #1.25
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/10/124,557
; ; FILING DATE: 16-Apr-2002
; ; CLASSIFICATION: <Unknown>
; ; PRIOR APPLICATION DATA:
; ; APPLICATION NUMBER: US 07/643,502
; ; FILING DATE: 18-JAN-1991
; ; APPLICATION NUMBER: US 07/546,114
; ; FILING DATE: 29-JUN-1990
; ; APPLICATION NUMBER: US 07/457,196
; ; FILING DATE: 29-DEC-1989
; ; APPLICATION NUMBER: US 07/390,901
; ; FILING DATE: 08-AUG-1989
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Geert, Luann
; ; REGISTRATION NUMBER: 31,822
; ; REFERENCE/DOCKET NUMBER: GI 5190
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: (617)876-1170
; ; TELEFAX: (617)876-5851
; ; INFORMATION FOR SEQ ID NO: 49:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 3945 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: double
; ; TOPOLOGY: unknown
; ; MOLECULE TYPE: cDNA
; ; FEATURE:
; ; NAME/KEY: CDS
; ; LOCATION: 1..3942
; ; SEQUENCE DESCRIPTION: SEQ ID NO: 49:
US-10-124-557-49

Alignment Scores:
Pred. No.: 0.000452 Length: 3945
Score: 144.00 Matches: 53
Percent Similarity: 35.14% Conservative: 12
Best Local Similarity: 28.65% Mismatches: 99
Query Match: 14.47% Indels: 22
DB: 13 Gaps: 4

US-10-030-225-2_COPY_27_213 (1-187) x US-10-124-557-49 (1-3945)

Qy 6 ProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAlaGluValSerAlaSer 25
Db 1634 CCACCCCTGAGAAGCCCGCACCCACCCCTGAG-GAGCTCGCACCCACCCCTGAG 1692
Qy 26 ProAspProArgProLeuLysGluGluGluAlaProLeuLeuProArgThrHisLeu 45
Db 1693 ---GAGCCACACCCACCCCTGAGGAGCTGCTCCACACTCCCAAGCAGCGGCT 1749
Qy 46 GlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluProAlaAlaMetThr--- 64
Db 1750 CCCAACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCT 1809
Qy 65 -----ProGlyAsnThrThrProArgThrProGluValThrProLeuArgLeuGlu 82
Db 1810 AAGGAGCCTGCTCCAACTACCCCTAAGGAGCTGCTCCAACTACCCCT----- 1857
Qy 83 LeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAspThrGln 102
Db 1858 -----AAGGGAGCTGCTCCAACTACCCCTCAAGGAACCTGCACCCACTACTCCC 1905
Qy 103 AlaSerAlaSerPro-----AspProArgProLeuArg 113
```

Tue Feb 22 09:49:23 2005

```

Db      1906 AAGAAGCTGCCCCCAAGAGCTTGCACCCACCAAGAGGCCACATCCACCACC 1965
Qy      114 GluGluGluGluAlaArgLeuLeuProArgThrHisLeuGlnAlaGluLeuHisGlnHis 133
Db      1966 TCTGACAGCCCGTCCAACTACCCCTAAGGGGACTGCTCCAACTACCCCTAAGGAGCCT 2025
Qy      134 GlyCysTrpThrValThrGluProAlaAlaLeuThrProGlyAsnAlaThrProProArg 153
Db      2026 GCTCCAACTACCCCTAAGGAGCCTGCTCCAACTACCCCTAAGGGGACTGCTCCAACTACC 2085
Qy      154 ThrGlnGluValThrProLeuLeuGluLeuGlnLysLeuProGluLeuValHisAla 173
Db      2086 CTCAGGAACCTGCACCCCACTACTCCCAAGAGCCTGCCCCCAAGAGCTTGCACCCACC 2145
Qy      174 ThrLeuSerThrPro 178
Db      2146 ACCACCAAGGGGCC 2160

```

Search completed: February 22, 2005, 06:10:23  
Job time : 5491 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 22, 2005, 01:12:02 ; Search time 3107 Seconds  
(without alignments)  
2290.962 Million cell updates/sec

Title: US-10-030-225-2\_COPY\_27\_213

Perfect score: 995

Sequence: 1 LPVKPRLRGRPGSLRRLA.....PELVHATLSTPNPDNQVTIK 187

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool\_p/US10030225/runat\_18022005\_100355\_2354/app\_query.fasta\_1.327  
-DB=EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -NORM=scorepct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10030225 @CGN 1 1 5180 @runat\_18022005\_100355\_2354 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: gb\_est1:.\*  
2: gb\_est2:.\*  
3: gb\_hc:.\*  
4: gb\_est3:.\*  
5: gb\_est4:.\*  
6: gb\_est5:.\*  
7: gb\_est6:.\*  
8: gb\_gsa1:.\*  
9: gb\_gsa2:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	992	99.7	790	5	BX336289
2	992	99.7	826	5	BX387234
3	992	99.7	866	5	BX365974
4	992	99.7	867	5	BX403263
5	992	99.7	868	5	BX365998
6	992	99.7	872	5	BX344975
7	992	99.7	881	5	BX345009
8	992	99.7	900	5	BX398022
9	992	99.7	917	5	BX334584

10	992	99.7	936	5	BX357282
11	992	99.7	936	5	BX381620
12	992	99.7	938	5	BX397246
13	992	99.7	947	5	BX359737
14	992	99.7	960	5	BX381477
15	992	99.7	967	5	BX357315
16	992	99.7	968	5	BX356306
17	992	99.7	971	5	BX336623
18	992	99.7	973	5	BX379676
19	992	99.7	975	5	BX382082
20	992	99.7	978	5	BX336717
21	992	99.7	979	5	BX335053
22	992	99.7	979	5	BX360184
23	992	99.7	980	5	BX360878
24	992	99.7	982	5	BX359034
25	992	99.7	983	1	AL554329
26	992	99.7	983	5	BX358077
27	992	99.7	983	5	BX361532
28	992	99.7	985	5	BX458619
29	992	99.7	990	5	BX360443
30	992	99.7	994	5	BX339696
31	992	99.7	3009	3	CR749863
32	987	99.2	708	5	BX324767
33	987	99.2	882	5	BX387069
34	987	99.2	924	5	BX378610
35	987	99.2	955	5	BX343795
36	987	99.2	985	3	CNSLT1180
37	985	99.0	984	5	BX337459
38	983	98.8	782	5	BX360064
39	982	98.7	960	1	AL549276
40	980	98.5	913	5	BX366004
41	979	98.4	912	5	BX344130
42	978	98.3	870	5	BX387067
43	978	98.3	1113	5	BX439514
44	975	98.0	967	5	BX335752
45	974	97.9	982	5	BX358526

#### ALIGNMENTS

RESULT 1  
BX336289  
LOCUS BX336289 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
DEFINITION clone CSODI027YA01 5-PRIME, mRNA sequence.  
ACCESSION BX336289  
VERSION BX336289.2  
KEYWORDS GI:46274766  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 790)  
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On May 2, 2003 this sequence version replaced gi:30341500.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r  
There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
http://www.genoscope.cns.fr/cdna?s=CSODI027AA01Q1Plac=7348.r.  
Location/Qualifiers  
1. .790  
/organism="Homo sapiens"

FEATURES  
source

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOD1027YA01"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and Ecor V sites of the pCMVSPORT 6 vector. Library was normalized."

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 2,686-62 Length: 790
Score: 992.00 Matches: 186
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.47% Mismatches: 0
Query Match: 99.70% Indels: 0
DB: 5 Gaps: 0

```

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX336289 (1-790)

```

QY 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
DB 93 CTCCTCCGTTGAAGAACCGCGCTCCGCGACACACCGCTGGAGCTCACGAGGCTCGCA 152
QY 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
DB 153 GAGGTCTCAGCTCCCGAGTCTTAGGCTCTGAGGAGAGGAGGAGGAGGAGGAGGAGG 212
QY 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
DB 213 CCCAGAACCCACCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 272
QY 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
DB 273 GCAGGCATGACCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 332
QY 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrLeuSerThrProAsnProAsp 100
DB 333 CTGGAGCTGCAGAGCTCCGCGGATGGCCACACACCTTGAGTACCCCTAACCTTGAT 392
QY 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
DB 393 ACCGAGGCTTACGCTCCCGAGTCTTAGGCTCTGAGGAGAGGAGGAGGAGGAGGAGGAGG 452
QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
DB 453 CTCCTCCAGAACCCACCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 512
QY 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
DB 513 CCAGCAGCCCTGACCCCGAGGAGTGCACGCTCCCGAGGAGGAGGAGGAGGAGGAGGAGG 572
QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
DB 573 CTGCTGGAGCTGCAGAGCTGCAGAAATTGGTCCACGCAACCTTGAGTACCCCTAACCTT 632
QY 181 AspAsnGlnValThrIleLys 187
DB 633 GATAACCAAGTGACCATCAAG 653

```

## RESULT 2

```

BX387234
LOCUS BX387234 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CSOD1034YP24 5-PRIME, mRNA sequence.
ACCESSION BX387234
VERSION BX387234.2 GI:46571918
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

## REFERENCE

1. (bases 1 to 826)  
 Li, W.B., Gruber, C., Jesses, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On May 8, 2003 this sequence version replaced gi:30447511.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and Ecor V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 7348.r

There is a virtual cDNA representing this cluster. For more  
 information about this cluster and the virtual cDNA, see  
 http://www.genoscope.cns.fr/cdna?s=CSAA014ZD04RM1&e=7348.r.

## FEATURES

## source

```

1..826
Location/Qualifiers
    organism="Homo sapiens"
    mol_type="mRNA"
    db_xref="taxon:9606"
    clone="CSOD1034YP24"
    tissue_type="PLACENTA COT 25-NORMALIZED"
    clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
    notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and Ecor V sites of the pCMVSPORT 6 vector. Library was normalized."

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 2,836-62 Length: 826
Score: 992.00 Matches: 186
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.47% Mismatches: 0
Query Match: 99.70% Indels: 0
DB: 5 Gaps: 0

```

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX387234 (1-826)

```

QY 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
DB 83 CTCCTCCGTTGAAGAACCGCGCTCCGCGACACACCGCTGGAGCTCACGAGGCTCGCA 142
QY 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
DB 143 GAGGTCTCAGCTCCCGAGTCTTAGGCTCTGAGGAGAGGAGGAGGAGGAGGAGGAGG 202
QY 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
DB 203 CCCAGAACCCACCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 262
QY 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
DB 263 GCAGCAGTGCAGAGCTCCCGAGTCCCGAGTCCCGAGTCCCGAGTCCCGAGTCCCGAG 322
QY 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrLeuSerThrProAsnProAsp 100
DB 323 CTGGAGCTGCAGAGCTCCCGGATGGCCACACACCTTGAGTACCCCTAACCTTGAT 382
QY 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
DB 383 ACCGAGGCTTACGCTCCCGAGTCTTAGGCTCTGAGGAGAGGAGGAGGAGGAGGAGGAGG 442
QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
DB 443 CTCCTCCAGAACCCACCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 502
QY 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
DB 503 CCAGCAGCCCTGACCCCGAGGAGTGCACGCTCCCGAGGAGGAGGAGGAGGAGGAGGAGG 562

```

QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 |||||  
 Db 563 CTGCTGGAGCTGCAGAACTGGTCAGATGGTCCAGCAACCTTGAGTACCCCTAACCCCT 622  
 |||||  
 QY 181 AspAsnGlnValThrLleLys 187  
 |||||  
 Db 623 GATAACCAAGGTGACCATCAAG 643  
 |||||

RESULT 3  
 BX365974  
 LOCUS  
 DEFINITION BX365974 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0D1031YB04 5-PRIME, mRNA sequence.  
 ACCESSION BX365974  
 VERSION BX365974.2 GI:46288863  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 866)  
 AUTHORS Li W.B., Gruber C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On May 5, 2003 this sequence version replaced gi:30370960.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 7348.r

There is a virtual cDNA representing this cluster. For more  
 information about this cluster and the virtual cDNA, see  
 http://www.genoscope.cns.fr/cdna?s=CS2BAX12F12\_AX212H4\_1&c=7348.r.

FEATURES  
 source  
 1..866  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1031YB04"  
 /tissue\_type="PLACENTA"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 3e-62 Length: 866  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX365974 (1-866)

QY 1 LeuProValLysLeuProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 |||||  
 Db 87 CTCCTCCGTCAGAAAGCCGCGCTCCGCGACACACGCTGGAGCTCACCAGGCTCGCA 146  
 |||||

QY 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 |||||  
 Db 147 GAGGTCTAGCTCCCGAGATCTTAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 206  
 |||||

QY 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluPro 60  
 |||||

Db 207 CCCAGAACCCACCTGTCAGGCGAGAGCCACCAACATGGATGCTGCACTGTCACTGAGCCA 266  
 |||||

QY 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 |||||

Db 267 GCAGCCATGATCCCGAGCAACACCCCTCCAGGACCCAGAGGTTACTCCGTTGGCG 326  
 |||||

QY 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 |||||

Db 327 CTGGAGCTGCAGAAAGCTGCCGGATTGGCCAAACAACCTTGAGTACCCCTTAACCCCTGAT 386  
 |||||

QY 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 |||||

Db 387 ACCAGGCTTCAGGCTCCCGAGATCCCTAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAG 446  
 |||||

QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTTPThrValThrGlu 140  
 |||||

Db 447 CTCCCAGAACCCACCTGTCAGGCGAGAGTACCAACATGGATGTTGAGACTGTCACTGAG 506  
 |||||

QY 141 ProAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 |||||

Db 507 CCAGCAGCCCTGATCCCGAGGAATGCCACGCTCCAGGACCCAGGAGGTTACTCCCTTG 566  
 |||||

QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 |||||

Db 567 CTGCTGAGCTGCAGAAAGCTGCCAGAAATTGGTCCACGCAACCTTGAGTACCCCTAACCCCT 626  
 |||||

QY 181 AspAsnGlnValThrLleLys 187  
 |||||

Db 627 GATAACCAAGGTGACCATCAAG 647  
 |||||

RESULT 4  
 BX403263  
 LOCUS  
 DEFINITION BX403263 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0D1075YH10 5-PRIME, mRNA sequence.  
 ACCESSION BX403263  
 VERSION BX403263.2 GI:46846816  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 867)  
 AUTHORS Li W.B., Gruber C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On May 13, 2003 this sequence version replaced gi:30615118.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 7348.r

There is a virtual cDNA representing this cluster. For more  
 information about this cluster and the virtual cDNA, see  
 http://www.genoscope.cns.fr/cdna?s=CS2BAX14ZE09\_AX242H3\_1&c=7348.r.

FEATURES  
 source  
 1..867  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1075YH10"  
 /tissue\_type="PLACENTA"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized."

ORIGIN

Alignment Scores:  
 Pred. No.: 3,01e-62 Length: 867  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservativity: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX403263 (1-867)

```

Qy 1 LeuProValLysLeuProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db 106 CTCCCGCTGAAGACCGCGCTCCGCGGACACCGCTGGAGGCTCACGAGGCTCGCA 165

Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db 166 GAGGCTCTCAGCTCCCGAGATCTAGGCTCTGAAGGAAGAGGAGGAGGACCACTGCTC 225

Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 226 CCCAGAACCCCTCAGGAGGAGGACACCAACATGATGCTGGACTGTCACTGAGGCA 285

Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 286 GCAGCCATGACCCAGGCAACACACCCCTCCAGGACCCAGAGTTTACTCCGTTGGCG 345

Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100
Db 346 CTGGAGCTGCAGAGCTCGCGGATGCGCAACACCAACCTTGAGTACCCCTAACCCCTGAT 405

Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db 406 ACCCAGGCTTACGCTCCCGAGATCTAGGCTCTGAGGGAAGAGGAGGAGGACGACTG 465

Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
Db 466 CTCCCGAGACCCACCTCAGGAGGAGGACGAGCTACACCAACATGATGTTGACTGTCACTGAG 525

Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
Db 526 CCAGCAGGCTTACGCTCCCGAGGATGCGCAACATGTTGTTCCAGCAACCTTGAGTACCCCT 585

Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
Db 586 CTGCTGGAGCTCAGAGCTGCGAATGTTGTTCCAGCAACCTTGAGTACCCCTAACCCCT 645

Qy 181 AspAsnGlnValThrIleLys 187
Db 646 GATAACCAAGTGACCATCAAG 666
  
```

RESULT 5  
 BX365998 868 bp mRNA linear EST 08-APR-2004  
 LOCUS  
 DEFINITION  
 clone CS0D1074Y001 5-PRIME, mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 EST.  
 BX365998.2 GI:46306893

SOURCE  
 Homo sapiens (human)

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

1 (bases 1 to 868)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On May 5, 2003 this sequence version replaced gi:30376966.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see [http://www.genoscope.cns.fr/cdna?s=CS2BAX142D10\\_AX242B5\\_1&c=7348.r](http://www.genoscope.cns.fr/cdna?s=CS2BAX142D10_AX242B5_1&c=7348.r).

#### FEATURES

Location/Qualifiers  
 source  
 1..868

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="CS0D1074Y001"

/tissue\_type="PLACENTA COT 25-NORMALIZED"

/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/notes="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

#### ORIGIN

Alignment Scores:  
 Pred. No.: 3,01e-62 Length: 868  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservativity: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX365998 (1-868)

```

Qy 1 LeuProValLysLeuProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db 85 CTCCCGCTGAAGACCGCGCTCCGCGGACACCGCTGGAGGCTCACGAGGCTCGCA 144

Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db 145 GAGGCTCTCAGCTCCCGAGATCTAGGCTCTGAAGGAAGAGGAGGAGGACCACTGCTC 204

Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 205 CCCAGAACCCACCTCAGGAGGAGGACCAACCCCTCCAGGACCCAGAGTTTACTCCGTTGGCG 264

Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 265 GCAGCCATGACCCAGGCAACACACCCCTCCAGGACCCAGAGTTTACTCCGTTGGCG 324

Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100
Db 325 CTGGAGCTGCAGAGCTCGCGGATGCGCAACCAACCTTGAGTACCCCTAACCCCTGAT 384

Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db 385 ACCCAGGCTTACGCTCCCGAGATCTAGGCTCTGAGGGAAGAGGAGGAGGACGACTG 444

Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
Db 445 CTCCCGAGACCCACCTCAGGAGGAGGACCAACCAACATGATGTTGAGTGTCACTGAG 504

Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
Db 505 CCAGCAGGCTTACGCTCCCGAGGATGCGCAACATGTTGTTCCAGCAACCTTGAGTACCCCT 564

Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
Db 565 CTGCTGGAGCTCAGAGCTGCGAATGTTGTTCCAGCAACCTTGAGTACCCCTAACCCCT 624

Qy 181 AspAsnGlnValThrIleLys 187
Db 625 GATAACCAAGTGACCATCAAG 645
  
```

RESULT 6

BX344975  
 LOCUS BX344975 872 bp mRNA linear EST 08-APR-2004  
 DEFINITION BX344975 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0DI005YP06 5-PRIME, mRNA sequence.  
 ACCESSION BX344975  
 VERSION BX344975.2 GI:46282437  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 872)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On May 2, 2003 this sequence version replaced gi:30346327.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 7348.r  
 There is a virtual cDNA representing this cluster. For more  
 information about this cluster and the virtual cDNA, see  
 http://www.genoscope.cns.fr/cdna?s=CS2BAX122A01\_AX19ZH11\_1&c=7348.r

## FEATURES

source

Location/Qualifiers

1..872  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DI005YP06"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3 03e-62 Length: 872  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservativeness: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0  
 US-10-030-225-2\_COPY\_27\_213 (1-187) x BX344975 (1-872)  
 QY 1 LeuProValLysLeuProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 DB 91 CTCCTCGGTGAAGAGCCGCGCTCCGCGGACCCAGCCGCTGAGAGCTCACGAGGCTGCA 150  
 QY 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 DB 151 GAGGTCTCAGCTCCCGAGATCTAGGCTCTGAAGGAAGAGGAGGAGGACCACTGCTC 210  
 QY 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTTPThrValThrGluPro 60  
 DB 211 CCCAGAACCCCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 270  
 QY 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 DB 271 GCAGCATGATGACCCAGGCAACACCACTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 330  
 QY 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 DB 331 CTGGAGCTGCAGAGCTCCGCGGATTGGCCCAACACACACACTTGGTACCCCTAACCCCTGAT 390

QY 101 ThrGlnAlaSerAlaSerProArgProLeuArgGluGluGluAlaArgLeu 120  
 DB 391 ACCCAGGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGGAAGAGGAGGAGGAGGAGGAGG 450  
 QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTTPThrValThrGlu 140  
 DB 451 CTCCTCGGTGAAGAGCCGCGCTCCGCGGACCCAGCCGCTGAGAGCTCACGAGCTGCTG 510  
 QY 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 DB 511 CCAGAGCTTCAGCTCCCGAGATCTTAGGCTCTGAGGGAAGAGGAGGAGGAGGAGGAGG 570  
 QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 DB 571 CTGCTGAGCTGCAGAGCTGCCAGATTGTCACGCAACCTTGGTACCTTACCTTACCTT 630  
 QY 181 AsnGlnValThrLys 187  
 DB 631 GATAACCAAGGTGACCATCAAG 651

## RESULT 7

BX345009

LOCUS

DEFINITION

BX345009 881 bp mRNA linear EST 07-APR-2004  
 clone CS0DI060YF04 5-PRIME, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 2, 2003 this sequence version replaced gi:30336333.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 7348.r  
 There is a virtual cDNA representing this cluster. For more  
 information about this cluster and the virtual cDNA, see  
 http://www.genoscope.cns.fr/cdna?s=CS2BAX14ZE04\_AX232E7\_1&c=7348.r.

## FEATURES

source

Location/Qualifiers

1..881  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DI060YF04"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:

Pred. No.: 3 07e-62 Length: 881  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservativeness: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0

US-10-030-225-2\_copy\_27\_213 (1-187) x BX345009 (1-881)

```

Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db 92 CTCCTCCGTGAAGAGCCGCGCTCCGCGGACACCGGCTGGAGCCTCACGAGCTCGCA 151
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db 152 GAGGTCTCAGCTCCCGAGTCTTAGGCTCTGAAGGAAGAGGAGGACCACTGCTC 211
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 212 CCCAGAACCCACCTGCAGGAGGAGCCACCAACATGATGCTGACTGTCACTGAGCCA 271
Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 272 GCAGCCATGACCCAGGCAACACCACTCCAGGACCCAGAGGTTTACTCCGTGGCG 331
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100
Db 332 CTGGAGCTGCAGAGCTCCGGGATTGGCAACACCACTTGATGACCCCTAACCTGAT 391
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db 392 ACCCAGGCTTCAGCTCCCGAGTCTTAGGCTCTGAGGAAGAGGAGGAGGACGACTG 451
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
Db 452 CTCCCGCAGAACCCACCTGCAGGAGGAGTACACCAACATGATGTTGACTGTCACTGAG 511
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
Db 512 CCAGCAGCCCTGACCCAGGGAATGCCACGCTCCAGGACCCAGAGGTTTACTCCCTTG 571
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
Db 572 CTGTGGAGCTGCAGAGCTGCAGAAATTGGTCCACGCAACCTTGATGATACCCCTAACCCCT 631
Qy 181 AspAsnGlnValThrIleLys 187
Db 632 GATAACCAAGTGACCATCAAA 652

```

RESULT 8  
 BX398022  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1. (bases 1 to 900)  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On May 13, 2003 this sequence version replaced gi:30621466.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
 end enriched, double-strand cDNA was digested with Not I and cloned  
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
 was normalized. Library was constructed by Life Technologies, a  
 division of Invitrogen. This sequence belongs to sequence cluster  
 7348.r  
 There is a virtual cDNA representing this cluster. For more  
 information about this cluster and the virtual cDNA, see  
 http://www.genoscope.cns.fr/cdna?8=CS0D1051CD05QF1&c=7348.r.

## FEATURES

## source

1..900  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1051YH09"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:  
 Pred. No.: 3,15e-62 Length: 900  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX398022 (1-900)

```

Qy 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20
Db 91 CTCCTCCGTGAAGAGCCGCGCTCCGCGGACACCGGCTGGAGCCTCACGAGCTCGCA 150
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40
Db 151 GAGGTCTCAGCTCCCGAGTCTTAGGCTCTGAGGAAGAGGAGGAGGACCACTGCTC 210
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60
Db 211 CCCAGAACCCACCTGCAGGAGGAGCCACCAACATGATGCTGACTGTCACTGAGCCA 270
Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80
Db 271 GCAGCCATGACCCAGGCAACACCACTCCAGGACCCAGAGGTTTACTCCGTGGCG 330
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100
Db 331 CTGGAGCTGCAGAGCTCCGGGATTGGCAACACCACTTGATGACCCCTAACCTGAT 390
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120
Db 391 ACCCAGGCTTCAGCTCCCGAGTCTTAGGCTCTGAGGAAGAGGAGGAGGACGACTG 450
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140
Db 451 CTCCCGCAGAACCCACCTGCAGGAGGAGTACACCAACATGATGTTGACTGTCACTGAG 510
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160
Db 511 CCAGCAGCCCTGACCCAGGGAATGCCACGCTCCAGGACCCAGAGGTTTACTCCCTTG 570
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180
Db 571 CTGTGGAGCTGCAGAGCTGCAGAAATTGGTCCACGCAACCTTGATGATACCCCTAACCCCT 630
Qy 181 AspAsnGlnValThrIleLys 187
Db 631 GATAACCAAGTGACCATCAAG 651

```

RESULT 9  
 BX334584  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 917 bp mRNA linear EST 07-APR-2004  
 BX334584 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0D1004YJ21 5-PRIME, mRNA sequence.  
 BX334584  
 EST.  
 Homo sapiens (human)  
 Homo sapiens



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 917)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 1, 2003 this sequence version replaced gi:30310346.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
http://www.genoscope.cns.fr/cdna?s=CS0D1004CE11QPI&c=7348.r.

FEATURES

source

1. 917  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1004YJ21"  
/tissue type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V sites of the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 3.23e-62 Length: 917  
Score: 992.00 Matches: 186  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.47% Mismatches: 0  
Query Match: 99.70% Indels: 0  
DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX334584 (1-917)

Oy 1 LeuProVallylsProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
Db 91 CTCCCGGTGAGAGCCGCGCTCCGCGACACCGCTGGAGCCTCAGAGCTCGCA 150  
Oy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluAlaProLeuLeu 40  
Db 151 GAGGTCTCAGCTCCCGAGATCCTAGGCCTCTGAAGGAAGAGGAGGACCACTGCTC 210  
Oy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
Db 211 CCCAGAACCCACCTGCGAGGAGCCACCAACATGGATGCTGACTGTCTACCTGAGCCA 270  
Oy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
Db 271 GCAGCCATGATCCCGAGGACCAACACCCCTCCAGGACCCAGAGGTTACTCCGTGGGG 330  
Oy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAspProAsp 100  
Db 331 CTGGAGCTGCAGAGCTGCGGGATTGGCCAAACACCACTTGTAGTACCCCTTAACCCCTGAT 390  
Oy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
Db 391 ACCAGGCTTCAGCTCCCGAGATCCTAGGCCTCTGAGGGAAGAGGAGGAGGACGACTG 450  
Oy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
Db 451 CTCCCCGAGAACCCACCTGCGAGGAGACTACCAACATGGATGTTGACTGTCTACTGTAG 510  
Oy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160

Db 511 CCAGCAGCCTGACCCCGAGGAATGCCAGCTCCAGGACCCAGGAGTTACTCCCTTG 570  
Oy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
Db 571 CTGCTGAGCTGCAGAGCTGCCAGAAATTGTCACGCAACCTTGAGTACCCCTAACCCCT 630  
Oy 181 AspAsnGlnValThrLeuLys 187  
Db 631 GATAACCCAGGTGACCATCAAG 651

RESULT 10

BX357282

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
7348.r

There is a virtual cDNA representing this cluster. For more  
information about this cluster and the virtual cDNA, see  
http://www.genoscope.cns.fr/cdna?s=CS0D1023CG12QPI&c=7348.r.

FEATURES

source

1. 936  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1023YN23"  
/tissue type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V sites of the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 3.31e-62 Length: 936  
Score: 992.00 Matches: 186  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.47% Mismatches: 0  
Query Match: 99.70% Indels: 0  
DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX357282 (1-936)

Oy 1 LeuProVallylsProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20

Db 87 CTCCCCGTGAAGAGCCGCGCTCCGCGACACCGCTGGAGCCTCAGAGCTCGCA 146

Oy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40

Db 147 GAGGTCTCAGCTCCCGAGATCCTAGGCCTCTGAGGGAAGAGGAGGAGGACCACTGCTC 206

Oy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60



1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?8=CS0D1032BE03QP1&c=7348.r>.

## FEATURES

source  
1. .938  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1032YI06"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:  
Pred. No.: 3.32e-62 Length: 938  
Score: 992.00 Matches: 186  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.47% Mismatches: 0  
Query Match: 99.70% Indels: 0  
DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX397246 (1-938)

Qy 1 LeuProVallyLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
Db 72 CTCCCGGTGAAGACCGCGCTCGGGACACCGCTGGAGCCTCACGAGCTCGCA 131  
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
Db 132 GAGTCTCAGCTCCCGAGATCCTAGGCTCTGAAGGAAGAGGAGGACCACTGCTC 191  
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTprThrValThrGluPro 60  
Db 192 CCCAGAACCCACCTGCGAGGACGACCAACCAACATGATGCTGACTGTCACTGAGCCA 251  
Qy 61 AlaAlaMetThrProGlyAsnThrThrProArgThrProGluValThrProLeuArg 80  
Db 252 GCAGCCATGACCCAGGCAACACCCCTCCAGGACCCAGAGTTACTCCGTTGGG 311  
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
Db 312 CTGGAGCTGCAGAGCTGCGGGATTGGCCAAACACCTTGAGTACCCCTAACCTGAT 371  
Qy 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
Db 372 ACCAGGCTTCAGCTCCCGAGATCCTAGGCTCTGAGGGAAGAGGAGGACGACGCTG 431  
Qy 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTprThrValThrGlu 140  
Db 432 CTCCCGAAGACCCACCTGCGAGGACGACCTACCAACATGATGTTGACTGTCACTGAG 491  
Qy 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
Db 492 CCAGCAGCCCTGACCCAGGGAATGCCAGCCTCCAGGACCCAGGAGTTACTCCCTTG 551  
Qy 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
Db 552 CTGCTGGAGCTGCAGAGCTGCCAGAAATTGTTCCACGCAACCTTGAGTACCCCTAACCC 611  
Qy 181 AspAsnGlnValThrLeuLys 187  
Db 612 GATAACCAAGGTGACCAATCAAG 632

## RESULT 13

BX359737 947 bp mRNA linear EST 08-APR-2004  
LOCUS : BX359737 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
DEFINITION : clone CS0D1060YF04 5-PRIME, mRNA sequence.  
ACCESSION : BX359737  
VERSION : BX359737.2 GI:46305712  
KEYWORDS : EST.  
SOURCE : Homo sapiens (human)  
ORGANISM : Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 947)  
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On May 5, 2003 this sequence version replaced gi:30376366.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more information about this cluster and the virtual cDNA, see <http://www.genoscope.cns.fr/cdna?8=CS0D1060DC02QP1&c=7348.r>.

## FEATURES

source  
1. .947  
/location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1060YF04"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Alignment Scores:  
Pred. No.: 3.36e-62 Length: 947  
Score: 992.00 Matches: 186  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.47% Mismatches: 0  
Query Match: 99.70% Indels: 0  
DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX359737 (1-947)

Qy 1 LeuProVallyLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
Db 92 CTCCCGGTGAAGACCGCGCTCGGGACACCGCTGGAGCCTCACGAGCTCGCA 151  
Qy 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
Db 152 GAGTCTCAGCTCCCGAGATCCTAGGCTCTGAAGGAAGAGGAGGACCACTGCTC 211  
Qy 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTprThrValThrGluPro 60  
Db 212 CCCAGAACCCACCTGCGAGGACGACCAACCAACATGATGCTGACTGTCACTGAGCCA 271  
Qy 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
Db 272 GCAGCCATGATCCCGAGCAACACCCCTCCAGGACCCAGAGTTACTCCGTTGGG 331  
Qy 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
Db 332 CTGGAGCTGCAGAGCTGCCGGGATTGGCCAAACACCTTGAGTACCCCTAACCCCTGAT 391

QY 101 ThrGlnAlaSerProAspProArgProLeuArgGluGluAlaArgLeu 120  
 Db 392 ACCAGGCTTCAGCTCCCGAGTCTAGGCTCTGAGGAGAGAGAGGACGACTG 451  
 QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
 Db 452 CTCCCGAGAACCCACCTGAGGAGAGAGTACACCAACATGGATGTTGGACTGCTCACTGAG 511  
 QY 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 Db 512 CCAGCAGCCCTGACCCCGAGGAATGCCAGCTCCCGAGGACCCAGGAGTTACTCCCTTG 571  
 QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 Db 572 CTGTGGAGCTGCAGAGCTGCAGAAATGGTTCACGCAACCTTGAGTACCCCTAACCTT 631  
 QY 181 AspAsnGlnValThrIleLys 187  
 Db 632 GATAACCAAGTGACCATCAAG 652

RESULT 14  
 BX381477  
 LOCUS  
 DEFINITION BX381477 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0DI066YK17 5-PRIME, mRNA sequence.

ACCESSION BX381477  
 VERSION  
 KEYWORDS  
 SOURCE EST. GI:46622135

ORGANISM Homo sapiens (human)

REFERENCE

AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On May 8, 2003 this sequence version replaced gi:30452951.  
 Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7348.r

There is a virtual cDNA representing this cluster. For more

information about this cluster and the virtual cDNA, see

http://www.genoscope.cns.fr/cdna?s=CS0DI066AF09QPI&c=7348.r.

Location/Qualifiers

FEATURES

source

1..960  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DI066YK17"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:  
 Pred. No.: 3.41e-62 Length: 960  
 Score: 992.00 Matches: 186  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.47% Mismatches: 0  
 Query Match: 99.70% Indels: 0  
 DB: 5 Gaps: 0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX381477 (1-960)

QY 1 LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla 20  
 Db 92 CTCCCGGTGAAGAGCCGCGCTCCGCGACCCAGCGCTGGAGCCTCACAGAGCTCGCA 151  
 QY 21 GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu 40  
 Db 152 GAGGTCTCAGCTCCCGAGTCTTAGGCTCTGAAGAGAGAGGAGGACCACTGCTC 211  
 QY 41 ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro 60  
 Db 212 CCCAGAACCCACCTGCAGGAGAGCCACCAACATGGATGCTGACTGTCACTGAGCCA 271  
 QY 61 AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg 80  
 Db 272 GCAGCCCATGATCCCGAGGCAACACCCCTCCAGGACCCCGAGGAGTTACTCCGTTGCGG 331

QY 81 LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrThrLeuSerThrProAsnProAsp 100  
 Db 332 CTGGAGCTGCAGAGCTGCCGGGATTGCCCAACAACCTTGAGTACCCCTAACCTTGAT 391

QY 101 ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu 120  
 Db 392 ACCAGGCTTCAGGCTCCCGAGTCTTAGGCTCTGAGGAGAGAGGAGGACGACTG 451

QY 121 LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu 140  
 Db 452 CTCCCGAGAACCCACCTGCAGGAGAGTACCAACATGGATGTTGAGTGTCACTGAG 511

QY 141 ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu 160  
 Db 512 CCAGCAGCCCTGACCCCGAGGAATGCCAGCTCCCGAGGACCCAGGAGTTACTCCCTTG 571

QY 161 LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro 180  
 Db 572 CTGTGGAGCTGCAGAGCTGCCAGAAATGGTTCACGCAACCTTGAGTACCCCTAACCTT 631

QY 181 AspAsnGlnValThrIleLys 187

Db 632 GATAACCAAGTGACCATCAAG 652

RESULT 15

BX357315

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 5, 2003 this sequence version replaced gi:30374142.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized. Library

was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

7348.r

There is a virtual cDNA representing this cluster. For more

information about this cluster and the virtual cDNA, see

http://www.genoscope.cns.fr/cdna?s=CS0DI024AD11QPI&c=7348.r.

Location/Qualifiers

FEATURES

source 1. .967  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI024YG21"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and Ecor V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

## Alignment Scores:

Pred. No.:	3.45e-62	Length:	967
Score:	992.00	Matches:	186
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.47%	Mismatches:	0
Query Match:	99.70%	Indels:	0
DB:	5	Gaps:	0

US-10-030-225-2\_COPY\_27\_213 (1-187) x BX357315 (1-967)

Qy	1	LeuProValLysLysProArgLeuArgGlyProArgProGlySerLeuThrArgLeuAla	20
Db	92	CTCCCGGTGAAGAAGCCGCGCTCGCGGACACCGGCTGGAGCCTCACGAGGCTCGCA	151
Qy	21	GluValSerAlaSerProAspProArgProLeuLysGluGluGluAlaProLeuLeu	40
Db	152	GAGGTCTCAGGCTCCCCAGATCTTAGGCCTCTGAGGAGAGAGGAGGACCACTGCTC	211
Qy	41	ProArgThrHisLeuGlnAlaGluProHisGlnHisGlyCysTrpThrValThrGluPro	60
Db	212	CCAGAACCCACCTGCAGGCAGAGCCACCAACATGATGCTGACTGTCACTGAGCCA	271
Qy	61	AlaAlaMetThrProGlyAsnThrThrProProArgThrProGluValThrProLeuArg	80
Db	272	GCAGCCATGACCCAGGCAACACCCCTCCAGGACCCAGAGGTTACTCCGTTGCGG	331
Qy	81	LeuGluLeuGlnLysLeuProGlyLeuAlaSerThrLeuSerThrProAsnProAsp	100
Db	332	CTGGAGCTGCARAAGCTCCCGGATTTGCCCAACACACCTTGATACCCCTAACCCCTGAT	391
Qy	101	ThrGlnAlaSerAlaSerProAspProArgProLeuArgGluGluGluAlaArgLeu	120
Db	392	ACCCAGGCTTCAGCTCCCGATCTTAGGCTCTGAGGAGAGAGAGGAGGACGACTG	451
Qy	121	LeuProArgThrHisLeuGlnAlaGluLeuHisGlnHisGlyCysTrpThrValThrGlu	140
Db	452	CTCCCGAGAACCCACCTGCAGGCAGAGCTACCAACATGGATTTGACTGTCACTGAG	511
Qy	141	ProAlaAlaLeuThrProGlyAsnAlaThrProProArgThrGlnGluValThrProLeu	160
Db	512	CCAGAGCCCTTGACCCAGGGAATGCCACGCTCCAGAGCCCGAGGAGTTACTCCCTTG	571
Qy	161	LeuLeuGluLeuGlnLysLeuProGluLeuValHisAlaThrLeuSerThrProAsnPro	180
Db	572	CTGCTGGAGCTGCAGAGCTGCCAGATTGTTCCACGCAACCTTGAGTACCCCTAACCCCT	631
Qy	181	AspAsnGlnValThrIleLys	187
Db	632	GATAACCAAGTGACCATCAAG	652

Search completed: February 22, 2005, 03:12:19  
Job time : 3111 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2005, 20:55:10 ; Search time 2878 Seconds  
(without alignments)  
10758.472 Million cell updates/sec

Title: US-10-030-225-1\_COPY\_58\_696  
Perfect score: 639  
Sequence: 1 atgcgtgcgtccgcgaccg.....ataaccaggtgacatcaag 639

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	639	100.0	2981	6	AX136231 Sequence
2	639	100.0	2981	6	BD093285 Different
3	639	100.0	2981	6	BD123568 Secretary
4	639	100.0	2981	9	AK075445 Homo sapi
5	635.8	99.5	2883	9	AJ583024 Homo sapi
6	542.6	84.9	750	6	AX136531 Sequence
7	542.6	84.9	750	6	BD123771 Secretary
8	522.4	81.8	3052	9	AK056709 Homo sapi
9	497.4	77.8	1698	6	C0720015 Sequence
10	475.2	74.4	2783	9	BC033140 Homo sapi
11	370	57.9	231464	9	AF111168 Homo sapi
12	261.8	41.0	497	6	CQ049341 Sequence
13	261.8	41.0	497	6	CQ064365 Sequence
14	261.8	41.0	497	6	CQ091307 Sequence
15	261.8	41.0	497	6	C0130141 Sequence
16	261.8	41.0	497	6	CQ168762 Sequence
17	261.8	41.0	497	6	CQ197880 Sequence
18	261.8	41.0	497	6	CQ213333 Sequence
19	261.8	41.0	497	6	CQ251912 Sequence

C 20	261.8	41.0	497	6	CQ289061 Sequence
C 21	261.8	41.0	497	6	CQ326064 Sequence
C 22	239.8	37.5	243	6	CQ054407 Sequence
C 23	239.8	37.5	243	6	CQ073647 Sequence
C 24	239.8	37.5	243	6	CQ104532 Sequence
C 25	239.8	37.5	243	6	CQ143244 Sequence
C 26	239.8	37.5	243	6	CQ178734 Sequence
C 27	239.8	37.5	243	6	CQ203077 Sequence
C 28	239.8	37.5	243	6	CQ226434 Sequence
C 29	239.8	37.5	243	6	CQ264575 Sequence
C 30	239.8	37.5	243	6	CQ301666 Sequence
C 31	239.8	37.5	243	6	CQ338879 Sequence
C 32	143	22.4	801	9	HS4330362 Homo sapi
C 33	125.2	19.6	172662	10	AC120540 Mus muscu
C 34	55.2	8.6	125020	9	AF429315 Homo sapi
C 35	51	8.0	125020	9	AF429315 Homo sapi
C 36	46.4	7.3	11687	1	AE004662 Pseudomon
C 37	44.8	7.0	2704	6	CQ843353 Sequence
C 38	44.8	7.0	2704	9	AK126288 Homo sapi
C 39	44.8	7.0	112027	9	AC073502 Homo sapi
C 40	44.4	6.9	110000	1	BX571965 33 o
C 41	44.4	6.9	110000	1	CP000010 24
C 42	43.2	6.8	1116	6	BD180389 Highly th
C 43	43.2	6.8	273285	1	AE017304 Thermus t
C 44	43	6.7	3177	9	AF055989 Homo sapi
C 45	43	6.7	110477	4	CR456637 Platypus

ALIGNMENTS

RESULT 1  
AX136231  
LOCUS AX136231 2981 bp DNA linear PAT 30-MAY-2001  
DEFINITION Sequence 153 from Patent EP1067182.  
ACCESSION AX136231  
VERSION AX136231.1 GI:14272639  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and  
Hayaishi, K.  
TITLE Secretory protein or membrane protein  
JOURNAL Patent: Ep 1067182-A 153 10-JAN-2001;  
Helix Research Institute (JP)  
FEATURES  
source  
1..2981  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
58..1773  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39758.1"  
/db\_xref="GI:14272640"  
/translation="MRALRDAGLLLCVLLAALLAALGLGVKVKPLRGRPGSLTR  
LAVVASPDPRPLKEEEAFLPRLTHQAEPHQHCWTVTEPAAMTFGNTTTPPTPEV  
TPRLLEQLPLGLASTLSTPNPTQASAPDRPLREBEERALLPRLTHQAELHQHG  
CAWVTEPAALPGNATPPRTQEVTPPLLEQLKPLVHATLSTPNPQNTVKVVEDP  
QWTSIDLLAEAPNPPQDTLSMLPALMPWLDGDKGEEDKAPGKGEKEEDVY  
SEIDEGEDQEDKEDEEQAALWNGTTDNDQGLARGDVFKDSVSYDPEQKWS  
WSPGNCSTGKQRTPCGYGTATRTCDLPCSGTEKDTGLGSPSEWKLARN  
ATDMHQDVDSCKNLNKSDFLIKYSQMLRDLPCPCAYLAMPDPSVSLQEHQ  
RSFRWRDASGRRLDIYQPTARFLRSLSGESSTLAQHCYDEDSRLTRKGAG  
MPLNLSTDFSPKLHFKPDTTPWILCKGWSRLHAVLPNNNGRACTDNPBEEYLAQLQ  
EAXEY"

ORIGIN

Query Match 100.0%; Score 639; DB 6; Length 2981;  
Best Local Similarity 100.0%; Pred. No. 6.5e-129;

Matches	639;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	ATCGTGTGCGTCTCGCGACCGAGCGGGCTCTCTCTCTGCGTGTCTGTCTGTGCGCGCGCTG	60						
Db	58	ATCGTGTGCGTCTCGCGACCGAGCGGGCTCTCTCTCTGCGTGTCTGTCTGTGCGCGCGCTG	117						
Qy	61	CTGAGGCGCGCTTAGGGCTCCCGTGAAGAAGCCGGGTCCCGGACACAGCGCTTGGG	120						
Db	118	CTGAGGCGCGCTTAGGGCTCCCGTGAAGAAGCCGGGTCCCGGACACAGCGCTTGGG	177						
Qy	121	AGCTCTACGAGGCTCGCAGAGGTCTCAGCTTCCCAGATCCTAGGCTCTCTGAAGGAAG	180						
Db	178	AGCTCTACGAGGCTCGCAGAGGTCTCAGCTTCCCAGATCCTAGGCTCTCTGAAGGAAG	237						
Qy	181	GAGGAGCACACTGTCTCCCAGAACCCACTGCGAGCAGAGCCACACCAATGGATGC	240						
Db	238	GAGGAGCACCACTGTCTCCCAGAACCCACTGCGAGCAGAGCCACACCAATGGATGC	297						
Qy	241	TGGACTGTCTACTGAGCCAGCAGCATACACCCAGGCAACACCCCTCCAGAGACCCCA	300						
Db	298	TGGACTGTCTACTGAGCCAGCAGCATACACCCAGGCAACACCCCTCCAGAGACCCCA	357						
Qy	301	GAGTTACTCCGTTGCGGCTGGAGCTGCAGAAGCTGCGGGATTGGCCAGCAACCTTG	360						
Db	358	GAGTTACTCCGTTGCGGCTGGAGCTGCAGAAGCTGCGGGATTGGCCAGCAACCTTG	417						
Qy	361	AGTACCCCTAACCCCTGATATCCAGGCTTCAGCCTCCCGAGATCCTAGGCTCTGAGGGAA	420						
Db	418	AGTACCCCTAACCCCTGATATCCAGGCTTCAGCCTCCCGAGATCCTAGGCTCTGAGGGAA	477						
Qy	421	GAGGAGGAGCAGACTGTCTCCCGAGAACCCACTGCGAGGAGAGCTACACCAACATGGA	480						
Db	478	GAGGAGGAGCAGACTGTCTCCCGAGAACCCACTGCGAGGAGAGCTACACCAACATGGA	537						
Qy	481	TGTTGGACTGTCACTGAGCCAGCAGCCCTGACCCGAGGGAATGCCAGCCTTCCAGGACC	540						
Db	538	TGTTGGACTGTCACTGAGCCAGCAGCCCTGACCCGAGGGAATGCCAGCCTTCCAGGACC	597						
Qy	541	CAGGAGGTACTCCCTTGTCTGTGAGCTGCAGAAGCTGCCAGAAATTTGTCACGCAACC	600						
Db	598	CAGGAGGTACTCCCTTGTCTGTGAGCTGCAGAAGCTGCCAGAAATTTGTCACGCAACC	657						
Qy	601	TTGAGTACCCCTAACCCCTGATAACCAAGGTGACCATCAAG	639						
Db	658	TTGAGTACCCCTAACCCCTGATAACCAAGGTGACCATCAAG	696						

RESULT 2	BD093285	2981 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	BD093285				
DEFINITION	Differentiation growth factor.				
ACCESSION	BD093285				
VERSION	BD093285.1	GI:22638873			
KEYWORDS	WO 01043312-A/1.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2981)				
AUTHORS	Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Yoshida, K. and Masuho, Y.				
TITLE	Differentiation growth factor				
JOURNAL	Patent: WO 0104312-A 1 18-JAN-2001; HELIX RESEARCH INSTITUTE, TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, KENJI YOSHIDA, YASUHIKO MASUHO				
COMMENT	OS Homo sapiens (human) PN WO 0104312-A/1 PD 18-JAN-2001 PF 06-JUL-2000 WO 2000JP004514 PR 08-JUL-1999 JP 99P 194179, 18-OCT-1999 US 60/159586 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, KENJI YOSHIDA, YOSHIDA, PI YASUHIKO MASUHO PC C12N15/16, C12N15/12, C12N15/85, C12N5/10, C12P21/02, C07K14/575,				





```
Db 238 GAGGAGGACCACTGCTCCAGAACCCACCTGCAGGAGAGCCACCAACATGGATGC 297
Qy 241 TGGAGTGTCACTGAGCAGCAGCATGACCCAGCAGACACACCCCTCCAGGAGCCCA 300
Db 298 TGGAGTGTCACTGAGCAGCAGCATGACCCAGCAGACACACCCCTCCAGGAGCCCA 357
Qy 301 GAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAGCACCAACTTG 360
Db 358 GAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAGCACCAACTTG 417
Qy 361 AGTACCCCTAACCTTGATACCAAGCTTCAGCTCCCGAGTCCCGAGTCTCAGGGAA 420
Db 418 AGTACCCCTAACCTTGATACCAAGCTTCAGCTCCCGAGTCTCAGGGAA 477
Qy 421 GAGGAGGAGCAGCAGCTCCCGAGACCCACCTGCAGGAGAGCTACACCAACATGGA 480
Db 478 GAGGAGGAGCAGCAGCTCCCGAGACCCACCTGCAGGAGAGCTACACCAACATGGA 537
Qy 481 TGTGGAGTGTCACTGAGCAGCAGCTCCCGAGACCCACCTGCAGGAGAGCTACACCAACATGGA 540
Db 538 TGTGGAGTGTCACTGAGCAGCAGCTCCCGAGACCCACCTGCAGGAGAGCTACACCAACATGGA 597
Qy 541 CAGGAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAGCACCAACTTG 600
Db 598 CAGGAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAGCACCAACTTG 657
Qy 601 TTGAGTACCCCTAACCTTGATACCAAGCTTCAGCTCCCGAGTCTCAGGGAA 639
Db 658 TTGAGTACCCCTAACCTTGATACCAAGCTTCAGCTCCCGAGTCTCAGGGAA 696
```

## RESULT 5

```
AJ583024
LOCUS AJ583024 2883 bp mRNA linear PRI 18-JUN-2004
DEFINITION Homo sapiens mRNA for Tail1 protein.
ACCESSION AJ583024
VERSION AJ583024.1 GI:48958174
KEYWORDS Tail1 gene; Tail1 protein; thrombospondin and AMOP containing
isthmin-like 1 protein.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Rossi V., Belfagna G., Rampazzo A., Baucce B. and Danieli G.A.
TITLE Tail1: an isthmin-like gene, containing type 1
thrombospondin-repeat and AMOP domain, mapped to ARVD1 critical
region
JOURNAL Gene 335, 101-108 (2004)
PUBMED 15194193
REFERENCE 2 (bases 1 to 2883)
AUTHORS Rossi V.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2003) Rossi V., Department of Biology, University
of Padua, viale G. Colombo, 3 - Padova, 35131, ITALY
FEATURES
source
1..2883
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3"
/tissue_type="heart"
/dev_stages="adult"
1..2883
/gene="TAIL1"
/gene="TAIL1"
/function="putative role in development"
/codon_start=1
/product="thrombospondin and AMOP containing isthmin-like
1 protein, (Tail1)"
```

## gene

## CDS

```
/protein_id="CAE47313.1"
/db_xref="GI:48958175"
/translation="MRALRDRAGLLLCVLLAALLAALGLPVKPKRLRGRPRGSLTR
LAEVSASDPRLPKKEEEAELLPRTHLQAEPHOHGWTVEPAAMTGNATPPRTPEV
TPRLLEQLPLGLANTLSTPNPDQTASAPDRPLREBEERLLPRTHLQAEHLQHG
CWITPEPALTPGNATPPRTOEVTPLLELQKLPYVHATLSTPNPDQVTTIKVEDP
QAEVSDLLAEPSPNPQDTLSWLPALWGDYKGEKDRAPFGEKGEKEBEDYV
SEDIEGSDQDEEDQALWFGNTDNDQGWLAFGDWVFKDSYDVEPQKWS
WSPCSGNCSTGKQKTRPCGCTATETRTCDLPSCPGCTKDTGLGPSBEWKLLAR
ATMDHDDVDSCEKLNCKSDFLKYLSOMLRDLPSCPGCAVPLEAMDSPVSLDEHOG
RSFRWRDASPRERLDIYQPTARFCLRSLMSGBSSTLAAOHCYDEDSRLLTRKGAG
MPNLISDFSPKLHFKFDTPWILLCKGDSRLHVLPPNNNGRACCTNPLBEEYLQQLQ
EAKFY"
36..113
/gene="TAIL1"
114..1748
/gene="TAIL1"
/product="Tail1 protein"
2865..2870
/gene="TAIL1"
```

## ORIGIN

```
Query Match 99.5%; Score 635.8; DB 9; Length 2883;
Best Local Similarity 99.7%; Pred. No. 3.2e-128;
Matches 637; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGCGTCCGCTCCGCGACCGAGCGCGGCTCTCTCTCTGCGTGTGCTGCTGGCGGCGCTG 60
Db 36 ATGCGTCCGCTCCGCGACCGAGCGCGGCTCTCTCTCTGCGTGTGCTGCTGGCGGCGCTG 95
Qy 61 CTGGAGCGCGGCTAGGGCTCCCGTGGAAGACCGCGGCTCCGCGGACCAACCGCCCTGGG 120
Db 96 CTGGAGCGCGGCTAGGGCTCCCGTGGAAGACCGCGGCTCCGCGGACCAACCGCCCTGGG 155
Qy 121 AGCCTCAGAGGCTCGCAGAGGCTTCAGCCTCCCGAGATCTTAGGCTCTGAAGAAGAG 180
Db 156 AGCCTCAGAGGCTCGCAGAGGCTTCAGCCTCCCGAGATCTTAGGCTCTGAAGAAGAG 215
Qy 181 GAGGAGCAGCAGCTGCTCCCGAGAACCCACCTGCGAGCAGAGCCACCAACATGGATGC 240
Db 216 GAGGAGCAGCAGCTGCTCCCGAGAACCCACCTGCGAGCAGAGCCACCAACATGGATGC 275
Qy 241 TGGAGTGTCACTGAGCAGCAGCCATGACCCCGAGCAACACCCCTCCAGGAGCCCA 300
Db 276 TGGAGTGTCACTGAGCAGCAGCCATGACCCCGAGCAACACCCCTCCAGGAGCCCA 335
Qy 301 GAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAGCACCAACTTG 360
Db 336 GAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAGCACCAACTTG 395
Qy 361 AGTACCCCTAACCTTGATACCCAGGCTTCAGCCTCCCGAGATCTTAGGCTCTCAGGGAA 420
Db 396 AGTACCCCTAACCTTGATACCCAGGCTTCAGCCTCCCGAGATCTTAGGCTCTCAGGGAA 455
Qy 421 GAGGAGGAGCAGCAGCTGCTCCCGAGAACCCACCTGCGAGCAGAGCTACCAACATGGA 480
Db 456 GAGGAGGAGCAGCAGCTGCTCCCGAGAACCCACCTGCGAGCAGAGCTACCAACATGGA 515
Qy 481 TGTGGAGTGTCTGCTGAGCAGCAGCCCTGACCCCGAGGAGTGCAGCCTCCAGGAGCC 540
Db 516 TGTGGAGTGTCTGCTGAGCAGCAGCCCTGACCCCGAGGAGTGCAGCCTCCAGGAGCC 575
Qy 541 CAGGAGGTTACTCCCTTGTGCTGGAGCTGCAGAGCTGCCAGAAATTTGGTCCAGCAACC 600
Db 576 CAGGAGGTTACTCCCTTGTGCTGGAGCTGCAGAGCTGCCAGAAATTTGGTCCAGCAACC 635
Qy 601 TTGAGTACCCCTAACCTTGATACCAAGCTTCAGCTCCCGAGTCTCAGGGAA 639
Db 636 TTGAGTACCCCTAACCTTGATACCAAGCTTCAGCTCCCGAGTCTCAGGGAA 674
```

## RESULT 6

AX136531



```

Qy 479 GATGTTGACTGTCTCACTGAGCAGCAGCCCTGACCCAGGAATGCCACGCTCCC-AGG 537
Db 538 GATGTTGACTGTCTCACTGAGCAGCAGCCCTGACCCAGGAATGCCACGCTCCCAGG 597
Qy 538 ACCCAGGAGGTTACTCCCTTCTGCTGAGCTGCA--GAAGCTGCCAG-AAATGGTCCAC 594
Db 598 ACCCAAGANGTTACTCCCTTCTGCTGAGCTGCAAGAAAGCTGCCAGAAATGGTCCAC 657
Qy 595 GCAA--CCTTGAGTACCCCTAACCTGATACACAGG 628
Db 658 GCAAACCTTGAGTTANCCCTAACCTGATAACCAAG 693

RESULT 8
AK056709 3052 bp mRNA linear PRI 30-JAN-2004
LOCUS Homo sapiens cDNA FLJ32147 f1s, clone PLACE5000116.
DEFINITION AK056709
ACCESSION AK056709
VERSION 1 GI:16552189
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayaashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Ohayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
Shiratori,A., Sudo,H., Hosiiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikkawa,E., Omura,Y., Abe,K., Kamiyama,K., Katsuta,N., Sato,K.,
Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yanashita,H.,
Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
Hara,H., Tanase,T., Nomura,Y., Togita,S., Komai,F., Hara,R.,
Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T.,
Shohata,N., Sano,S., Moriya,S., Momiya,H., Satoh,N., Takami,S.,
Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
Kunagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hiki,I., Kobatake,N.,
Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
Naguchi,S., Itoh,T., Shigetake,K., Senba,T., Matsumura,K.,
Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039
2
Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and
Isogai,T.
NEBO human cDNA sequencing project
Unpublished
3 (bases 1 to 3052)
Isogai,T., Otsuki,T. and Sugiyama,T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,

```

```

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEBO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.
FEATURES             Location/Qualifiers
     source            1. 3052
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="PLACE5000116"
                        /tissue_type="placenta"
                        /clone_lib="PLACE5"
                        /note="cloning vector: pME18SFL3"
ORIGIN
Query Match      81.8%; Score 522.4; DB 9; Length 3052;
Best Local Similarity 85.8%; Pred. No. 1.6e-103;
Matches 638; Conservative 0; Mismatches 1; Indels 105; Gaps 1;
Qy 1 ATGCGTGGCTCCCGCAGACCGAGCGGGCTCTCTCTCTGCGTGTCTGCTGGCGCGCTG 60
Db 24 ATGCGTGGCTCCCGCAGACCGAGCGGGCTCTCTCTCTGCGTGTCTGCTGGCGCGCTG 83
Qy 61 CTGAGAGCGCGCTGAGGGCTCCCGTGAAGAGCGCGGCTCCCGGACACACGCGCTGGG 120
Db 84 CTGAGAGCGCGCTGAGGGCTCCCGTGAAGAGCGCGGCTCCCGGACACACGCGCTGGG 143
Qy 121 AGCTCTCAGGGCTCGAG----- 139
Db 144 AGCTCTCAGGGCTCGAGAGCGCGCTTTTCAGGTGCGCGCTAGCTCGGAGAGAGTCCC 203
Qy 140 ----- 139
Db 204 GCGGTGGAGGAATCTGGCGCTTTTGGAGTGGCGGCGCGCGCTGAGGTTCAAAC 263
Qy 140 ----AGTCTCAGGCTCCCGAGATCTTAGGCTCTGAAGAGAGAGAGGACCACTG 195
Db 264 GGTGAGTCTCAGGCTCCCGAGATCTTAGGCTCTGAAGAGAGAGAGGACCACTG 323
Qy 196 CTCCCCAGACCCACCTGCGAGGAGAGCCACCAACATGATGCTGCTGCTGCTGCTGAG 255
Db 324 CTCCCCAGACCCACCTGCGAGGAGAGCCACCAACATGATGCTGCTGCTGCTGCTGAG 383
Qy 256 CCAGCAGCCATGACCCCGCAGCAACACCCCTCCCGAGGACCCCGAGAGGTTTACTCCGTTG 315
Db 384 CCAGCAGCCATGACCCCGCAGCAACACCCCTCCCGAGGACCCCGAGAGGTTTACTCCGTTG 443
Qy 316 CGGTGAGCTGCGAGAGCTGCGGGATGTCGAGCAACCACTTGTAGTACCCCTAACCT 375
Db 444 CGGTGAGCTGCGAGAGCTGCGGGATGTCGAGCAACCACTTGTAGTACCCCTAACCT 503
Qy 376 GATACCCAGGCTTCAGGCTCCCGAGATCTTAGGCTCTGAGGGAAGAGAGGAGGACGCA 435
Db 504 GATACCCAGGCTTCAGGCTCCCGAGATCTTAGGCTCTGAGGGAAGAGAGGAGGACGCA 563
Qy 436 CTGCTCCCGAGAACCCACCTGCGAGGAGAGCTACACCAACATGGATGTGGACTGTCACT 495
Db 564 CTGCTCCCGAGAACCCACCTGCGAGGAGAGCTACACCAACATGGATGTGGACTGTCACT 623
Qy 496 GAGCAGCAGCCCTGACCCCGAGGAATGCCAGCGCTCCCGAGGACCCAGAGGTTTACTCCC 555
Db 624 GAGCAGCAGCCCTGACCCCGAGGAATGCCAGCGCTCCCGAGGACCCAGAGGTTTACTCCC 683
Qy 556 TTGCTGCTGGAGCTGCAGAGAGCTGCCAGAAATTTGTCACCGAACTTGTAGTACCCCTAAC 615
Db 684 TTGCTGCTGGAGCTGCAGAGAGCTGCCAGAAATTTGTCACCGAACTTGTAGTACCCCTAAC 743

```

QY 616 CCTGATACCAAGGTGACCATCAAG 639  
 Db 744 CCTGATACCAAGGTGACCATCAAG 767

RESULT 9  
 CQ720015

LOCUS CQ720015 1698 bp DNA linear PAT 03-FEB-2004  
 DEFINITION Sequence 5949 from Patent WO02068579.  
 ACCESSION CQ720015  
 VERSION CQ720015.1 GI:42280872  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.  
 Kits, such as nucleic acid arrays, comprising a majority of  
 human exons or transcripts, for detecting expression and other uses  
 thereof  
 JOURNAL Patent: WO 02068579-A 5949 06-SEP-2002;  
 PE Corporation (NY) (US)  
 FEATURES source  
 Location/Qualifiers  
 1..1698  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

ORIGIN  
 Query Match 77.8%; Score 497.4; DB 6; Length 1698;  
 Best Local Similarity 99.8%; Pred. No. 4.6e-98;  
 Matches 498; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 141 GGTCTCAGCTCCCGAGTCTTAGGCTCTGAAGAGAGAGAGGAGGACACACTGCTCC 200  
 Db 402 GGTCTCAGCTCCCGAGTCTTAGGCTCTGAAGAGAGAGAGGAGGACACACTGCTCC 461  
 QY 201 CAGAACCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 260  
 Db 462 CAGAACCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 521  
 QY 261 AGCATGACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 320  
 Db 522 AGCATGACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 581  
 QY 321 GGAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 380  
 Db 582 GGAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 641  
 QY 381 CCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGG 440  
 Db 642 CCAGGCTTACGCTCCCGAGATCTTAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGG 701  
 QY 441 CCCAGAGACCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 500  
 Db 702 CCCAGAGACCCACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 761  
 QY 501 AGCAGCCCTGACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 560  
 Db 762 AGCAGCCCTGACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 821  
 QY 561 GCTGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 620  
 Db 822 GCTGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 881  
 QY 621 TAACCAAGGTGACCATCAAG 639  
 Db 882 TAACCAAGGTGACCATCAAG 900

RESULT 10  
 BC033140

LOCUS BC033140 2783 bp mRNA linear PRI 19-JUL-2004  
 DEFINITION Homo sapiens thrombospondin, type I, domain containing 3, mRNA  
 (CDNA clone IMAGE:3844836), partial cds.  
 ACCESSION BC033140  
 VERSION BC033140.1 GI:21619879  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2783)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Aitschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haie, F.,  
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettner, M., Madan, A., Young, A.C., Shvachenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smaluk, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 PUBMED 12477932  
 REFERENCE 2 (bases 1 to 2783)  
 Strausberg, R.  
 Direct Submission  
 Submitted (25-JUN-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK  
 NTH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgapbs-x@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcgsc.bc.ca  
 Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,  
 Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth  
 Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,  
 Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,  
 Teika Olson, Diana Palmquist, Anca Petrescu, Anna Lissa Prabhu,  
 Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,  
 Duane Smaluk, Jeff Stott, Miranda Tsai, George Yang, Jacquie  
 Schein, Asim Siddiqui, Rob Holt, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAL Plate: 43 Row: c Column: 10.  
 FEATURES source  
 Location/Qualifiers  
 1..2783  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3844836"  
 /tissue type="Placenta, choriocarcinoma"  
 /clone\_lib="NIH MGC 21"  
 /lab\_host="DH10B-R"  
 /notes="Vector: pOT7"  
 <1..2783  
 /gene="THSD3"





```
QY 442 CCCAGAACCCACCTGCAGGACAGCTACACCAACATGGATGTTGACTGTCACTGAGCCA 501
Db 87103 CCCAGAACCCACCTGCAGGACAGCTACACCAACATGGATGTTGACTGTCACTGAGCCA 87044
QY 502 GCAGCCCTGACCCAGGGAATGCCACGCTCCAGAGCCAGGAGGTTACTCCCTTGTG 561
Db 87043 GCAGCCCTGACCCAGGGAATGCCACGCTCCAGAGCCAGGAGGTTACTCCCTTGTG 86984
QY 562 CTGGAGCTGCAGAGCTGCCAGAAATGTTCCACGCAACCTTGAGTACCCCTAACCTGAT 621
Db 86983 CTGGAGCTGCAGAGCTGCCAGAAATGTTCCACGCAACCTTGAGTACCCCTAACCTGAT 86924
QY 622 AACCAAGTGA 631
Db 86923 AACCAAGTGA 86914

RESULT 12
LOCUS CQ049341/c
DEFINITION Sequence 161 from Patent WO0157270.
ACCESSION CQ049341
VERSION CQ049341.1 GI:41023818
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.
TITLE Human genome-derived single exon nucleic acid probes useful for
analysis of gene expression in human breast and hbl 100 cells
JOURNAL Patent: WO 0157270-A 161 09-AUG-2001;
Aeomica, Inc. (US)
FEATURES
source
1. .497
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="MAP TO AF111168.2-EXPRESSED IN HBL100, SIGNAL =
0.82"

ORIGIN
Query Match 41.0%; Score 261.8; DB 6; Length 497;
Best Local Similarity 77.3%; Pred. No. 1e-46; Indels 11; Gaps 1;
Matches 384; Conservative 0; Mismatches 2;
QY 171 GAAGGAAGAGGAGGAGGACCACTGCTCCCCAGAACCCACCTGCAGGAGGAGCCACACCA 230
Db 497 GAAGGAAGAGGAGGAGGACCACTGCTCCCCAGAACCCACCTGCAGGAGGAGCCACACCA 438
QY 231 ACATGATGCTGACTGTCACTGAGCCAGAGCCATGACCCAGGCAACACACCCCTCC 290
Db 437 ACATGATGCTGACTGTCACTGAGCCAGAGCCATGACCCAGGCAACACACCCCTCC 378
QY 291 CAGGACCCAGAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAG 350
Db 377 CAGGACCCAGAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAA 318
QY 351 CACAACTTGAGTACCCCTAACCTGTATACC----- 381
Db 317 CACAACTTGAGTACCCCTAACCTGTATACCAGGTGAGAGCTACAGAAGGCCAGCAGC 258
QY 382 ----- 381
Db 257 TCCTGGCTCTGCTGCGCCAGCCTTTTCTGGGTCCCTGTTGGGTAAGCCCTAACCTGTAC 198
QY 382 -----CAGGCTTCAGCTCCCCAGATCCTAGGCTCTTGAGGGA 419
Db 197 CTCACCTTCCCTCTCTCTTCTCAGGCTTACGCTCCCGATCCTAGGCTCTTGAGGGA 138
QY 420 AGAGGAGGAGGACGAGTGTCTCCAGAACCCACCTGCAGGAGAGCTTACACCAACATGG 479
Db 137 AGAGGAGGAGGACGAGTGTCTCCAGAACCCACCTGCAGGAGAGCTTACACCAACATGG 78
QY 480 ATGTTGACTGTCTACTGAGCCAGGAGCCCTGACCCAGGGAATGCCAGGCTCCAGGAC 539
Db 77 ATGTTGACTGTCTACTGAGCCAGGAGCCCTGACCCAGGGAATGCCAGGCTCCAGGAC 18
```

```
Db 137 AGAGGAGGAGGACGAGTGTCTCCAGAACCCACCTGCAGGAGAGCTTACACCAACATGG 78
QY 480 ATGTTGACTGTCTACTGAGCCAGGAGCCCTGACCCAGGGAATGCCAGGCTCCAGGAC 539
Db 77 ATGTTGACTGTCTACTGAGCCAGGAGCCCTGACCCAGGGAATGCCAGGCTCCAGGAC 18
QY 540 CCAGGAGTGTACTCCCT 556
Db 17 CCAGGAGTGTACTCCCT 1

RESULT 13
LOCUS CQ064365/c
DEFINITION Sequence 165 from Patent WO0157278.
ACCESSION CQ064365
VERSION CQ064365.1 GI:41034232
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.
TITLE Human genome-derived single exon nucleic acid probes useful for
analysis of gene expression in human hela cells or other human
cervical epithelial cells
JOURNAL Patent: WO 0157278-A 165 09-AUG-2001;
Aeomica, Inc. (US)
FEATURES
source
1. .497
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="MAP TO AF111168.2-EXPRESSED IN HELA, SIGNAL = 0.97"

ORIGIN
Query Match 41.0%; Score 261.8; DB 6; Length 497;
Best Local Similarity 77.3%; Pred. No. 1e-46; Indels 11; Gaps 1;
Matches 384; Conservative 0; Mismatches 2;
QY 171 GAAGGAAGAGGAGGAGGACCACTGCTCCCCAGAACCCACCTGCAGGAGGAGCCACACCA 230
Db 497 GAAGGAAGAGGAGGAGGACCACTGCTCCCCAGAACCCACCTGCAGGAGGAGCCACACCA 438
QY 231 ACATGATGCTGACTGTCACTGAGCCAGAGCCATGACCCAGGCAACACACCCCTCC 290
Db 437 ACATGATGCTGACTGTCACTGAGCCAGAGCCATGACCCAGGCAACACACCCCTCC 378
QY 291 CAGGACCCAGAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAG 350
Db 377 CAGGACCCAGAGGTTACTCCGTTGCGGCTGGAGCTGCAGAGCTGCCGGGATGGCCAA 318
QY 351 CACAACTTGAGTACCCCTAACCTGTATACC----- 381
Db 317 CACAACTTGAGTACCCCTAACCTGTATACCAGGTGAGAGCTACAGAAGGCCAGCAGC 258
QY 382 ----- 381
Db 257 TCCTGGCTCTGCTGCGCCAGCCTTTTCTGGGTCCCTGTTGGGTAAGCCCTAACCTGTAC 198
QY 382 -----CAGGCTTCAGCTCCCCAGATCCTAGGCTCTTGAGGGA 419
Db 197 CTCACCTTCCCTCTCTCTTCTCAGGCTTACGCTCCCGATCCTAGGCTCTTGAGGGA 138
QY 420 AGAGGAGGAGGACGAGTGTCTCCAGAACCCACCTGCAGGAGAGCTTACACCAACATGG 479
Db 137 AGAGGAGGAGGACGAGTGTCTCCAGAACCCACCTGCAGGAGAGCTTACACCAACATGG 78
QY 480 ATGTTGACTGTCTACTGAGCCAGGAGCCCTGACCCAGGGAATGCCAGGCTCCAGGAC 539
Db 77 ATGTTGACTGTCTACTGAGCCAGGAGCCCTGACCCAGGGAATGCCAGGCTCCAGGAC 18
```





**This Page Blank (uspto)**